Article

Big-Data Management: A Driver for Digital Transformation?

Panagiotis Kostakis [ID] and **Antonios Kargas** *[ID]

Department of Informatics & Telecommunications, University of Athens, 15784 Athens, Greece;
pakostak@di.uoa.gr
* Correspondence: akargas@di.uoa.gr

Abstract: The rapid evolution of technology has led to a global increase in data. Due to the large volume of data, a new characterization occurred in order to better describe the new situation, namel. big data. Living in the Era of Information, businesses are flooded with information through data processing. The digital age has pushed businesses towards finding a strategy to transform themselves in order to overtake market changes, successfully compete, and gain a competitive advantage. The aim of current paper is to extensively analyze the existing online literature to find the main (most valuable) components of big-data management according to researchers and the business community. Moreover, analysis was conducted to help readers in understanding how these components can be used from existing businesses during the process of digital transformation.

Keywords: big data; big-data management; digital transformation; Industry 4.0; information

1. Introduction

The Fourth Industrial Revolution or Industry 4.0 can be characterized as the Era of Digitalization and Information. Even though its initial aim was to deliver "fundamental improvements to the industrial process involved in manufacturing, engineering, material usage and supply chain and life cycle management" [1], soon enough it become clear that the boundaries of expected change would involve businesses in general and society as a whole [2].

Industry 4.0 aims to establish a constant interaction and communication among people, machines, and resources, at least in terms of exchanging data and information [3]. This involves the integration of various elements such as devices and machinery alongside physical elements (i.e., products and consumers) by using networked sensors and specialized software [4] leading to the decentralization of the operational decision-making process [5]. Such a procedure develops a complex system of high accuracy and speed that is capable to predict, plan, and control business outcomes [6], changing how both service providers and consumers think and act [7].

The "promise" of integration among physical objects, human actors, intelligent machines, production lines, and processes in order to develop a new agile, networked, and intelligent value chain [8] has involved business pioneers with a digital transformation procedure [9,10]. Digital transformation involved changing the nature and culture of product or service, the process of producing and delivering, production and business structures, operational management, and the practices used to control arising complexity [11].

Such a transformation aims to design and develop products or services (a) with embedded knowledge about consumers' preferences (information or data) [12], (b) capable of rearranging their characteristics according to these preferences (transforming information or data to business knowledge) [13] and to be produced or distributed with minimal human intervention while considering parameters such as tracked life cycle and customer use [14].

Collecting data on a mass scale and extracting knowledge out of them, which leads to automated production, leads to the need for big-data management [15,16]. By realizing that Industry 4.0 will gradually be applied to any business sector (where the Internet and embedded systems can be used) and to a growing number of everyday activities (directly

or indirectly related with consuming preferences and habits), led to increasing tension to explore the benefits of technologies such as big data management [17,18].

Moreover, since Industry 4.0 is an ongoing procedure, big-data management gained significant interest during the COVID-19 pandemic crisis [19,20], where industries, businesses, citizens, and governments were forced to face changes in their traditional operational models and to fasten their digital transformation [21,22]. Digital transformation is crucial in smaller countries, where financial, technological, and human resources are scarce, while gaining a competitive advantage on the European or global level is much more difficult [23,24].

The current paper aims to help readers in better understanding how big-data management is used in the existing literature by conducting quantitative text analysis with a qualitative literature review, considering that its meaning and usefulness vary across business sectors. The used methodology revealed big-data management's most significant components according to existing research, while these components can become the first priority for businesses targeting to achieve digital transformation. In Section 2, the concepts of big data, big-data management, and digital transformation are presented, and Section 3 describes the methodological approach. The proposed results are presented in Section 4, while Section 5 evaluates the proposed results regarding big-data management and their main components in terms of usefulness during a digital transformation process. Lastly, Section 6 outlines the most significant conclusions.

2. Literature Review

The term "big data" has gained much attention since the beginning of 21st century, with various researchers attempting to establish a widely accepted definition. One of the most common definitions was proposed by Gartner, who based it on a report of the META Group [25], which introduced the 3Vs that specify the big-data challenge: volume (vast amounts of data), velocity (fast data streams), and variety (heterogeneous content). It followed Roger Magoulas of O'Reilly Media in 2005 [26], who defined big data as a large volume of data, structured or unstructured, which traditional data-processing techniques are unable to manage and process due to the complexity and volume of the data.

Schroeck in 2012 [27] introduced a fourth V to the definition of big data, namely, veracity, with the overall definition comprising high velocity, large volume, wide variety, and uncertain veracity. A fifth V, namely, value, was added to the definition a year later [28], including the real value that the data can offer to the process or activity to which they are related after their processing. Of course, there were definitions from other researchers, emphasizing other aspects than the Vs, describing big data as information assets that, after processing, play a key role in the decision making and insight of a business [29].

Lastly, it was only in 2015 that the NIST Big Data Public Working Group standardized the proposed definition, linking big data with the 5Vs while emphasizing the need for efficient storage, manipulation, and analysis to reach a meaningful result [30]. The above-mentioned definition regards the 5Vs that characterize the quality of the big-data 5Vs [31,32]:

- Volume: the amount of stored and managed data.
- Velocity: the required computational speed to put a query in the data in relation to their rate of change.
- Variety: different forms of data (e.g., text, audio, and video).
- Veracity: the confidentiality of data.
- Value: the importance that organizations and entities attach to accessing data.

Since then, big data and their management has gained research interest, and they are applied to several sectors, including health [33], the environment [34], energy and energy management [35,36], geospatial applications [37], smart cities [38,39], real estate [40], and information sciences. Regarding information sciences, big-data management was applied to various research eras, such as cognitive computing [20,31–42], cloud computing [43,44], and cloud data management [45,46].

However, the vast amounts that big data represent have not complicated the processes in some other fields. For example, many e-commerce companies rely on their recom-

mender systems in order to process user ratings and preferences, and recommend the best items to sell to them. The high volume of data has caused scalability problems to surface, increasing the amount of processing time [47,48]. Even so, some implementations combining traditional algorithms and big-data technologies help in reducing the scalability problem [48,49].

It seems that data are omnipresent and overflowing, being almost everywhere in business and everyday life [50]. Yearly data collected from the Library of Congress are almost 235 terabytes [51], McKinsey estimated that Facebook's content exceeds 30 billion bits, while the value of data regarding the healthcare sector is more than USD 300 billion [52], while the International Data Corporation (IDC) predicted the expansion of digital universe from 4.4 ZB in 2003 to 44 ZB by 2020 [53]. For 2020, the global information capacity was estimated to reach 4 zettabytes (4 billion terabytes) [54], leading to a global explosion of used data, taking into consideration that the estimated number was forecast to double every 3 years.

In order to meet such forecasts and estimations, a series of advances took place in terms of tools, technologies, and operations. Big data require specialized tools in order to lead to significant results, which cannot be achieved with common methods and techniques. The focus of the processing is not so much their quantity, but the fact that they can add to the creation of information and knowledge, making companies more competitive and giving them the opportunity to offer better services to consumers and citizens.

In the following subsections, the terms "big data management" and "digital transformation" are analyzed in more detail.

2.1. Big-Data Management

The term "big-data management" refers to a set of data management practices. It is a mixture of old and new practices, skills, teams, data types, and functionality. With big data, businesses were forced to change, as it was difficult to manage vast amounts of data, and this change has led to the expansion of data management skills and software, and business process automation, bringing to the surface both technological and business issues. [55].

The transfer from the management of a traditional volume of data to the management of big data requires a change within the company [56]. There are five areas of interest:

- **Leadership**: The leadership team, which sets clear goals, determines success, and asks the right questions must also lead the company to an effective big data management system. Big data require the need for human guidance on the road to change and success. Evaluating information and extracting knowledge that can lead to successive business decisions is a science in itself, requiring visionary leadership.

- **Talent management**: The complexity and management of big data has to do both with the technology and the processes, and with scientific and professional personnel, the key persons whose job it is to implement, integrate, and keep operational such systems. The selection of specialized professionals and data scientists is necessary.

- **Technology**: Big-data management technology has constantly improved during the last few years. A series of tools have been developed for professional and scientific work, while open-source tools are available for the wide community of big-data management enthusiasms (e.g., Hadoop). So, IT departments have a variety of tools and solutions to integrate them with the rest of the organization's systems, but implementing and operating big data management systems also require significant skills that employees must acquire and constantly develop.

- **Decision making**: Information and decision making are inter-related elements in the everyday work and operational life cycle of an organization. Information is created and transferred within the organization through data processing. That is why it is important for people who manage and process data to work with people who are responsible for understanding the company's problems, finding solutions and making decisions.

- **Company culture**: A company's culture is shaped or reshaped by the way that data (and big data) are managed. Big data may lead a company nowhere, but transforming big data into valuable information and decision-making knowledge means a series of internal changes to organizational culture. Being sensitive to external environmental information (big data transformed into information) requires significant changes in terms of company culture.

2.2. Digital Transformation

The term "digital transformation" refers to the use of technology to gradually improve the performance of a business. Digital technologies and techniques such as analytics, mobility, social media, and smart devices are used with traditional technologies in order to change customer relationships, internal processes, and value propositions [57].

Before companies start to implement changes to move into the digital age, it is important to understand the logic of digitization and how digital transformation affects business. Figure 1 shows the drivers of digital transformation and the four levels in which it has an effect. These four levels are as follows.

- **Digital data**: acquiring, processing, and analyzing digital data leads to better forecasting and decision making.
- **Automation**: the integration of technology with artificial intelligence gives impetus to systems that autonomously work and are organized, leading to a reduction in errors and operating costs, and an increase in speed.
- **Connectivity**: the interconnection of all systems through high-bandwidth telecommunication networks synchronizes the supply chain and reduces production times.
- **Digital customer access**: Internet access gives businesses instant access to customers, providing them full transparency and new services.
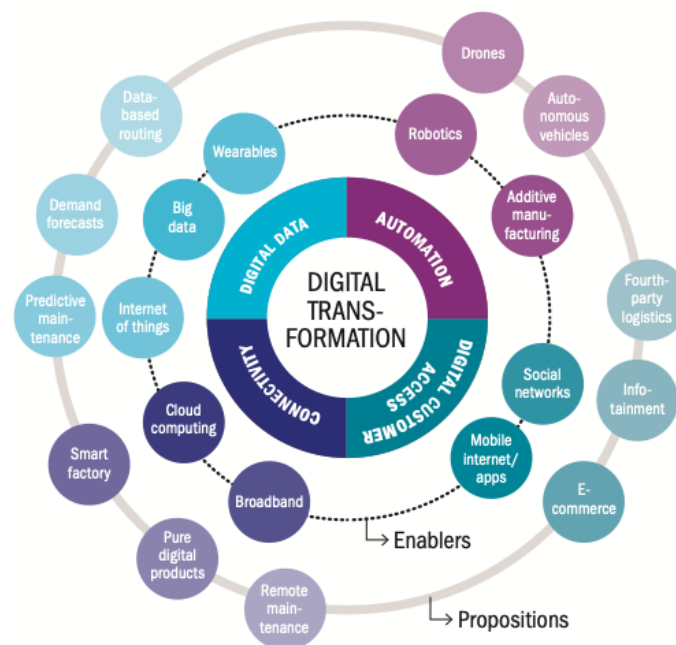

Figure 1. Digital transformation drivers, according to [58].

Figure 1 also depicts several assets that could help businesses in gaining access to those four levels, and eventually achieving digitalization and digital transformation.

The (a) availability of digital data, (b) process automation, (c) interconnection of production and supply chains, and (d) the creation of digital interfaces for customers as a whole are transforming business models and lead to business reorganization [58].

3. Methods

We identify key issues related with big-data management in the international literature. These issues may be related with both data gathering and manipulation, and the exploitation of the final result. The proposed research methodology was based on both (a) quantitative text analysis and (b) a qualitative literature review.

The former is mainly used to extract key words or even phrases from various sources, including documents on a large scale [59,60]. Big data management has garnered significant research interest during the last few years, leading to a constantly increasing number of publications. Even though quantitative analysis can provide significant results, combining it with qualitative analysis [61] can deepen our understanding about the importance of big-data management in the research community [62].

We conducted extensive research in Google Scholar's publication base, which is used as a repository for a large amount of research work. "Big-data management" was used as a key research term in titles, keywords, and abstracts [63,64]. The whole procedure revealed a total of 17,700 unique sources with big-data management in their analysis. More detailed analysis revealed that almost 1030 of them had big-data management in the core of their analysis, providing significant results, while the research interest of the rest, even though they included the term, targeted a different aspect (big-data management was only a supplementary element of their research).

Table 1 presents the percentage of papers including one of the presented terms in combination with the major research term of big-data management. Results indicate that the vast majority of papers relate big-data management with terms such as "information", "technology", and "business", while significant and extended usage exists when a paper's content is related with the energy or health industry. Even though the percentage for "digital transformation" seems small related with the above-mentioned terms, it is gaining attention among researchers, supporting authors' research interest on the current paper's topic.

Table 1. Paper content with the term "big-data management".

Content	Paper Percentage
Information	88%
Technology	76%
Business	64%
Energy Industry	37%
Health Industry	36%
Digital Transformation	4%

Quantitative text analysis was conducted to extract a list of unique words (unigrams) out of the proposed sources. Punctuation and capitalization were excluded while a series of words were simplified by removing their endings [65] in a procedure called word stemming. For example, the terms "analyze, analyse, analysis, analytics" were clustered under the term "analy"; following the same procedure, terms "decide, decision" were included in the term "deci". Moreover, meaningless words (e.g., adverbs and articles) and general terms (e.g., "presents" and "depicts") were excluded from further analysis.

By reducing complexity as described above, frequency analysis was conducted to count the number of occurrences of most significant words and phrases related with big-data-management articles. Lastly, following Grimmer and Steward [59] recommendations, we excluded less significant words and phrases (when one or two appearances occurred), and terms related to the words "big", "data", and "management" (separately or in combination).

Results indicated the existence of 142 significant words, and the 10 most significant words in terms of number of occurrences are presented in Table 2.

Table 2. Ten most significant words related to big-data management.

Word	Number of Occurrences
Analy	392
Tech	316
Info	286
System	282
Process	278
Stor	226
Company	192
App	154
Cloud	148
Deci	148

Table 3 presents the most significant phrases (bigrams and trigrams) revealed from the above-mentioned methodology in terms of the number of occurrences. Such an addition is useful in order to include significant terms such as "cloud computing" and "Internet of Things" that are often related to big-data management.

Table 3. Ten most significant phrases related to big-data management.

Phrase	Number of Occurrences
Data analytics	228
Decision making	108
Internet of Things	106
Data processing	104
Data storage	96
Information system	84
Cloud computing	80
Data mining	46
Social network	36
Supply chain	32

Qualitative literature review was used to evaluate the exact use of the above-mentioned significant terms in the scientific sources and documents. Understanding the exact use of the terms permitted the authors to develop seven clusters incorporating different areas of interest among researchers of big-data management [63,64]. The results of qualitative analysis are presented in Table 4, showing the proposed clusters and the three most significant words or phrases incorporating each cluster.

Table 4. Seven clusters of big-data management and three most significant words or phrases.

Clusters	Words/Phases	Number of Occurrences
I. Data Analysis	1. (Data) analy 2. (Data) process 3. Collect Total: 22	392 (228) 282 (104) 132 1588
II. Technology	1. Tech 2. System 3. App Total: 33	316 284 162 2238
III. Information and Knowledge	1. Info 2. Knowledge 3. Performance Total: 9	286 102 102 708
IV. Data Storage	1. (Data) stor 2. Volume 3. Architecture Total: 13	278 (96) 178 116 988
V. Business and Human Power	1. Company 2. Decision 3. Industry Total: 38	226 144 110 1488
VI. Data Type and Visualization	1. Model 2. Scalability 3. Complex Total: 11	102 86 80 608
VII. Security and Threats	1. Security 2. Privacy 3. Risk management Total: 12	60 52 24 232

After the initial clustering, further analysis was conducted regarding cluster content (in terms of big-data management) and possible interconnections between the various clusters. Results indicate that the seven above-mentioned clusters can be grouped into four major categories to reveal most significant aspects of big-data management tensions. The four proposed groups are:

1. **Data life-cycle processes**: Data Analysis (Cluster I), Data Storage (Cluster IV), and Data Type and Visualization (Cluster VI). All the above are strongly related with big data's life-cycle process and could be unified as a single management procedure.
2. **Technology** (Cluster II): remains as it is.
3. **Information Security**: including Information and Knowledge (Cluster III), and Security and Threats (Cluster VII). Extracting information and delivering knowledge from big data can be a competitive advantage in globalized economies, ensuring viability and growth in business environments. Under these conditions, the security of data reflects the ability of any company to protect its source of competitive advantage and to make valuable decisions while minimizing risks.
4. **Business and Human Power** (Cluster V): remains as it is.

Table 5 presents the four groups of principles incorporating the most significant aspects of big-data management.

Table 5. Main big-data management principles.

Groups	Total of Words/Phrases	Number of Occurrences
I. Data Life-Cycle Processes	Total: 46	3184
II. Technology	Total: 33	2238
III. Information Security	Total: 21	940
IV. Business and Human Power	Total: 38	1488

4. Results

Quantitative and qualitative text analysis led to the identification of four clusters that are the four main components of big-data management: Data Life-Cycle Processes, Technology, Information Security, and Business and Human Power.

These four components can each or all be identified in each publication concerning big-data management. Big data can be found in both the technological and business worlds, processes are actions using big data, and information comes from them. The following subsections analyze the four big-data management components.

4.1. Big-Data Life-Cycle Processes

Big-data life-cycle processes are actions and procedures that are executed in both technological and business environments. For example, processes such as data storage and processing are handled using technological tools, applications, and techniques, while processes such as data generation and their effective usage take place in a business environment.

However, that does not mean that there is a separation between the processes; on the contrary, technology and business cooperate, and their coexistence is inevitable. A significant example is the need for both technology and business for data analysis.

The data life cycle is the object of research and modeling [65,66]. In the first approach [65], the big data life cycle consist of 5 steps, as follows.

- **Acquiring data**: In this step, the source of data, their format, and where their extraction takes place are defined. In the case of a special type of format, then their storage is appropriately adapted, and the search and their formatting are rationalized.
- **Choosing architecture**: Because of the large amount of data that are processed, the architecture of the environment into which the data are inserted is important. The choice is made on the basis of cost and performance.
- **Shaping data**: before uploading them into the computing platform, data must also be in a suitable and compatible format.
- **Write code**: the right choice of programming language (e.g., R, Python) is also important and must be compatible with the system's technology (e.g., Hadoop).
- **Debugging and iteration**: the last step, in which results of data processing take a meaningful form and are visualized.

In the second approach [66], the steps of this model remind of the steps of solving a problem, seeing big data as a solution. The five steps of the model are:

- **Define the concern**: the problem that needs to be solved using big data is defined.
- **Search**: the big-data space is examined for data elements that could map the problem.
- **Transform**: the extract, transform, load (ETL) technique is used to extract data, transform them into appropriate formats, and store them for processing.
- **Entity resolution**: verification that the selected data elements are relevant and refer to the entity of problem.
- **Solve the problem**: preselected data are processed to compute the answer to the problem.

4.2. Technology

The big-data sector, as a technological sector, is close to technological changes and opportunities. There are a variety of technologies, tools, and techniques for collecting,

storing, processing, and analyzing big data. Some of these technologies were developed due to Big Data, while others pre-existed and evolved to be able to meet their specifications.

There are many techniques, technologies, and tools proposed for big-data management functions [67]. Among the most important techniques applied in the fields of statistics and computer science are:

- **Data mining**: technique of data pattern extraction from large volumes of data using statistical methods and machine learning.
- **Genetic algorithms**: Technique used for optimization, mainly for use in nonlinear problems.
- **Machine learning**: technique that uses the principles of artificial intelligence and through algorithms locates complex patterns in large volumes of data to make decisions.
- **Neural networks**: their practices are used to detect patterns in large volumes of data.

Among the most important proposed technologies and tools are [55,67]: Hadoop, MapReduce, Business Intelligence, Cloud Computing, Cloudera, Oracle Big Data Appliance, Pentaho, SAP Hana, Cassandra, MongoDB and Amazon Dynamo, with R, Java, and Python programming languages being the most popular.

4.3. Business and Human Power

Except for technology, another important element in the management of big data is the people who manage them. New skills need to be developed, and constant training is a business investment to future potentialities. However, there are limitations regarding both big-data detection and perception, leading to difficulties during the processes required for huge volumes of data [67].

The new skills are both for the people who manage big data and those who use, process, and manage the technologies, especially decision makers. Skill development is needed for every person that is directly or indirectly involved with the process of locating the information needed to make important decisions in this large volume of data.

New skills also create new roles and positions in organizational charts [55]. In addition to data scientists, organizations believe that data-architect, data-analyst, business-intelligence-manager, application-developer, business-analyst, and system-analyst or -architect positions should be available for big-data management.

The power and peculiarities of big data do not eliminate the need for a human point of view. The most important thing within an organization is to make the right decisions. Big-data management, on the other hand, makes it necessary to have individuals and teams who manage big data and make such decisions in order to gain a competitive advantage [56].

4.4. Information Security

Businesses and organizations that own big data process and analyze large amounts of data in order to extract meaningful information. Each organization has its own policy for the protection and security of its sensitive information. Protecting them is a major issue for big-data management, as there is a high security risk associated with big data [68], which is why information security is a major challenge for big data and their management.

Big-data security can be achieved using techniques such as authentication, authorization, and encryption. Various security measures that big-data applications face are network size, variety of devices, real-time security monitoring, and the lack of intrusion systems [69]. Therefore, great attention should be paid to the development of a multilevel security policy model and intrusion prevention systems.

Technologies such as cloud computing are complementary to big data in the field of information security. By improving system efficiency and providing additional cloud storage features, they can protect sensitive data and monitor access to them [69].

5. Discussion

Having entered the Fourth Industrial Revolution, also known as the Society of Information and Knowledge, companies need to change their strategies and practices in

order to cope with the information storm, rapid changes in the market with ultimate goals, competitive advantage and survival.

Big data, a technology that is gaining increasing attention, is a driving force in company goals to transform and gain a competitive advantage [58]. However, big data alone are not the driver that companies use in their strategies. They need to be properly managed, so that each pillar of big-data management to be integrated within a sufficient digital transformation's plan.

The technologies that have been developed with the arrival of big data offer more opportunities to businesses. Research conducted in companies showed that knowledge generated from data, and consequently from large volumes of data such as big data, leads to competitive advantages [70].

However, it is up to companies to choose the type of strategy that they follow [71]. With the revolutionary approach, all data are transferred to the new environment created with big data, and all processes are executed according to new models being developed [72]. On the other hand, in the evolutionary approach, big data are treated by companies in the same way and with the same means that they have for the management and processing of their data, gaining from the benefits they bring but having to face the problems that their management in old systems bring [72]. The third approach is the hybrid where, depending on the type of data, their management is chosen from either existing solutions or new technologies.

Whatever strategy is chosen or promoted, it definitely brings changes in business processes, data sources, infrastructure, architecture, skills, organizational structures, and the economy. In a business, the change and adoption of new strategies must come from the cooperation of both its leaders and IT, as the business and technological environment must be in harmony to properly manage and exploit the benefits of big data. In addition to technological solutions and changes [72], ideas and practices from people and the business itself as an entity should be included [73]. Such practices are:

- **Recognition of the uniqueness of big data**: The peculiarities of big data affect every part of the business, and the result they bring is uncertain. It is also uncertain whether they are able to realize the results they bring. That is why it is important to understand the principles and practices of big data.
- **Generation of new ideas**: in order for businesses and their leaders to transform, they must generate new ideas that provide answers to new questions and issues that arise.
- **Build business leadership belief in big data**: Leadership in various companies is not always willing to rely on results that bring data to make decisions, especially when they must create strategies. However, because information today is more complex, it is necessary to have faith in the results of big data.
- **Adoption of new investment plans**: Although the acquisition of big data does not greatly affect a company's finances, investment plans must be adjusted, so that the profits from big data exceed the costs of their overall management and processing.
- **Ensuring appropriate infrastructure**: it is also important for the IT department of a company to ensure that the organization could have the appropriate infrastructure, so that all big-data processes can be effortlessly executed.
- **Preparing for business risks**: Big data, in addition to benefits, also carry risks for business, especially since many of the data are often personal and sensitive. That is why they should be added to the strategic plans for their control and monitoring.
- **Expansion of existing skills**: understanding how big-data management and processing processes should be properly and efficiently performed requires a wealth of skills from people who already work or are to be hired in a company.
- **Change of organizational structures**: It is not always easy or even welcome by entities that comprise an organization to change. That is why there should be a plan to maximize the returns on investing in big data.

Of course, implementing large-scale changes alongside difficulties coming from the COVID-19 business environment is almost inevitable for most enterprises. Even before

COVID-19 occurred, implementation difficulties had been reported. Deloitte reported that few businesses reached large-scale changes, even though most of them believe that such change is inevitable [74]. A more recent study [75] examined the digital transformation of Japanese business, recognizing sources of pressure such as: (a) overseas partners, (b) fintech, (c) the drop in the Japanese population, (d) generational changes, (e) the preference for food grown in Japan, and (f) the upcoming innovation wage that would most probably impact all areas of life and business. The study asked whether the human resources and investments needed to bring digital transformation to fruition exist.

The importance of big data to the whole procedure of digital transformation was highlighted by the EU Commission, which proposed four main categories: mobile, social media, cloud, and data analytics [76].

6. Conclusions

In the era of big data, businesses must learn the true value and benefits of data. They must change strategies in order to elicit the whole range of information and knowledge from processing large volumes of data, such as big data, so that to ensure viability and to gain a competitive advantage under uncertain conditions.

The current paper contributes to the discussions around topics of big-data management and digital transformation, which are garnering interest from the business and scientific worlds. Quantitative and qualitative text analysis led to the identification of four main components of big-data management, namely, data life-cycle processes, technology, information security, and business and human power. The proposed analysis found and clarified big data's most valuable components in terms of both technology and business operation. Moreover, these components of big-data management contribute to the identification, development, and implementation of ideas, tactics, and strategies necessary to digitally reform a business and develop its digital identity.

For each of the above-mentioned components, a wide bibliographic review was conducted in order to reveal possible business strategies that could lead or facilitate digital transformation. The proposed results indicate that the most valuable components of big-data management can provide a highway for digital transformation, while results seem to agree with various research.

The paper's contribution is as follows: (a) expanding the research on big-data management, helping to clarify its term and to understand the depth of its application and value; and (b) the research results offer helpful guidance in business strategy development, and more precisely on how to use big-data management practices to facilitate or achieve digital transformation.

References

1. Kagermann, H.; Wahlster, W.; Helbig, J.; Hellinger, A.; Stumpf, M.A.V.; Treugut, L.; Findeklee, U. Recommendations for Implementing the Strategic Initiative Industrie 4.0. In *Final Report of the Industrie 4.0. Frankfurt*; Forschungsunion: Berlin, Germany, 2013. Available online: http://alvarestech.com/temp/tcn/CyberPhysicalSystems-Industrial4-0.pdf (accessed on 23 August 2021).
2. Ebert, C.; Duarte, C. Requirements engineering for the digital transformation: Industry panel. In Proceedings of the Requirements Engineering Conference IEEE 24th International, Beijing, China, 12–16 September 2016; pp. 4–5.
3. Sima, V.; Gheorghe, I.G.; Subić, J.; Nancu, D. Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review. *Sustainability* **2020**, *12*, 4035. [CrossRef]

4. Shu, T.; Chuan, T.; Lee, A.; Ahmad, A.; Aizat, A.N. An Overview of Industry 4.0: Definition, Components, and Government Initiatives. *J. Adv. Res. Dyn. Control. Syst.* **2018**, *10*, 14.

5. Nagy, J.; Oláh, J.; Erdei, E.; Máté, D.; Popp, J. The Role and Impact of Industry 4.0 and the Internet of Things on the Business Strategy of the Value Chain—The Case of Hungary. *Sustainability* **2018**, *10*, 3491. [CrossRef]

6. Industrial Internet Consortium. Fact Sheet. 2013. Available online: https://www.iiconsortium.org/docs/IIC_FACT_SHEET.pdf (accessed on 2 March 2019).

7. Kargas, A.; Varoutas, D. Industry 4.0 in Cultural Industry. A Review on Digital Visualization for VR and AR Applications. In *Impact of Industry 4.0 on Architecture and Cultural Heritage*; Bolognesi, C.M., Cettina, S., Eds.; IGI Global: Hershey, PA, USA, 2020; pp. 1–19. [CrossRef]

8. Schumacher, A.; Erol, S.; Sihn, W.A. Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises. *Procedia CIRP* **2016**, *52*, 161–166. [CrossRef]

9. Vaidya, S.; Ambad, P.; Bhosle, S. Industry 4.0—A Glimpse. *Procedia Manuf.* **2018**, *20*, 233–238. [CrossRef]

10. Schwab, K. *The Fourth Industrial Revolution*; World Economic Forum: Geneva, Switzerland, 2017.

11. Matt, C.; Hess, T.; Benlian, A. Digital transformation strategies. *Bus. Inf. Syst. Eng.* **2015**, *57*, 339–343. [CrossRef]

12. Dwivedi, Y.; Ismagilova, E.; Hughes, L.; Carlson, J.; Filieri, R.; Jacobson, J.; Jain, V.; Karjaluoto, H.; Kefi, H.; Krishen, A.; et al. Setting the future of digital and social media marketing research: Perspectives and research propositions. *Int. J. Inf. Manag.* **2021**, *59*, 102168. [CrossRef]

13. Werner, R.; Wiegand, N.; Imschloss, M. The impact of digital transformation on the retailing value chain. *Int. J. Res. Mark.* **2019**, *36*, 350–366.

14. Hermann, M.; Pentek, T.; Otto, B. Design Principles for Industrie 4.0 Scenarios. In Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016; pp. 3928–3937. [CrossRef]

15. Gudivada, V.; Apon, A.; Ding, J. Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *Int. J. Adv. Softw.* **2017**, *10*, 1–20.

16. Pouyanfar, S.; Yang, Y.; Chen, S.; Shyu, M.; Iyengar, S. Multimedia Big Data Analytics: A Survey. *ACM Comput. Surv.* **2018**, *51*, 1–34. [CrossRef]

17. Fitzgerald, M.; Kruschwitz, N.; Bonnet, D.; Welch, M. Embracing Digital Technology: A New Strategic Imperative. *MIT Sloan Manag. Rev. Res. Rep.* **2013**, *55*, 1.

18. Ross, J.; Sebastian, I.; Beath, C.; Scantlebury, S.; Mocker, M.; Fonstad, N.; Kagan, M.; Moloney, K.; Geraghty Krusel, S. *Designing Digital Organizations*; MIT Center for IS Research: Cambridge, MA, USA, 2016.

19. Alsunaidi, S.J.; Almuhaideb, A.M.; Ibrahim, N.M.; Shaikh, F.S.; Alqudaihi, K.S.; Alhaidari, F.A.; Khan, I.U.; Aslam, N.; Alshahrani, M.S. Applications of Big Data Analytics to Control COVID-19 Pandemic. *Sensors* **2021**, *21*, 2282. [CrossRef] [PubMed]

20. Riswantini, D.; Nugraheni, E.; Arisan, A.; Khotimah, P.H.; Munandar, D.; Suwarningsih, W. Big Data Research in Fighting COVID-19: Contributions and Techniques. *Big Data Cogn. Comput.* **2021**, *5*, 30. [CrossRef]

21. Subramaniam, R.; Singh, S.P.; Padmanabhan, P.; Gulyás, B.; Palakkeel, P.; Sreedharan, R. Positive and Negative Impacts of COVID-19 in Digital Transformation. *Sustainability* **2021**, *13*, 9470. [CrossRef]

22. Kudyba, S. COVID-19 and the Acceleration of Digital Transformation and the Future of Work. *Inf. Syst. Manag.* **2020**, *37*, 284–287. [CrossRef]

23. Kargas, A.; Kiriakidis, M.; Zacharakis, E. Europe's Economic Crisis: Re–Clustering European Economies. *Eur. J. Soc. Sci. Educ. Res.* **2020**, *7*, 41. [CrossRef]

24. Laitsou, E.; Kargas, A.; Varoutas, D. Digital Competitiveness in the European Union Era: The Greek Case. *Economies* **2020**, *8*, 85. [CrossRef]

25. Laney, D. *3D Data Management: Controlling Data Volume, Velocity, and Variety*; META Group: Stamford, CT, USA, 2001.

26. Roger Magoulas on Big Data. Available online: http://radar.oreilly.com/2010/01/roger-magoulas-on-big-data.html (accessed on 21 December 2020).

27. Schroeck, M.; Shockley, R.; Smart, J.; Romero-Morales, D.; Tufano, P. *Analytics: The Real-World Use of Big Data—How Innovative Enterprises Extract Value from Uncertain Data*; IBM Institute for Business Value: New York, NY, USA, 2012.

28. Demchenko, Y.; Grosso, P.; de Laat, C.; Membrey, P. Addressing Big Data Issues in Scientific Data Infrastructure. In Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 20–24 May 2013; pp. 48–55.

29. Sicular, S. Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s. 2013. Available online: http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/#95a45853bf622013 (accessed on 10 August 2021).

30. Chang, W.L. *NIST Big Data Interoperability Framework: Volume 1, Definitions*; NIST Big Data Public Working Group: Gaithersburg, MD, USA, 2015.

31. Russom, P. Big data analytics. *TDWI Best Pract. Rep. Fourth Quart.* **2011**, *19*, 1–35.

32. Demchenko, Y.; Laat, C.; Membrey, P. Defining Architecture Components of the Big Data Ecosystem. In Proceedings of the International Conference on Collaboration Technologies and Systems, Minneapolis, MN, USA, 19–23 May 2014; pp. 104–112.

33. Štufi, M.; Bačić, B.; Stoimenov, L. Big Data Analytics and Processing Platform in Czech Republic Healthcare. *Appl. Sci.* **2020**, *10*, 1705. [CrossRef]

34. Dwevedi, R.; Krishna, V.; Kumar, A. Environment and Big Data: Role in Smart Cities of India. *Resources* **2018**, *7*, 64. [CrossRef]

35.   Marinakis, V.; Koutsellis, T.; Nikas, A.; Doukas, H. AI and Data Democratisation for Intelligent Energy Management. *Energies* **2021**, *14*, 4341. [CrossRef]

36.   Hernández-Moral, G.; Mulero-Palencia, S.; Serna-González, V.I.; Rodríguez-Alonso, C.; Sanz-Jimeno, R.; Marinakis, V.; Dimitropoulos, N.; Mylona, Z.; Antonucci, D.; Doukas, H. Big Data Value Chain: Multiple Perspectives for the Built Environment. *Energies* **2021**, *14*, 4624. [CrossRef]

37.   Li, Z.; Tang, W.; Huang, Q.; Shook, E.; Guan, Q. Introduction to Big Data Computing for Geospatial Applications. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 487. [CrossRef]

38.   Cerquitelli, T.; Migliorini, S.; Chiusano, S. Big Data Analytics for Smart Cities. *Electronics* **2021**, *10*, 1439. [CrossRef]

39.   Borrajo, L.; Cao, R. Big-But-Biased Data Analytics for Air Quality. *Electronics* **2020**, *9*, 1551. [CrossRef]

40.   Munawar, H.S.; Qayyum, S.; Ullah, F.; Sepasgozar, S. Big Data and Its Applications in Smart Real Estate and the Disaster Management Life Cycle: A Systematic Analysis. *Big Data Cogn. Comput.* **2020**, *4*, 4. [CrossRef]

41.   Hassani, H.; Huang, X.; MacFeely, S.; Entezarian, M.R. Big Data and the United Nations Sustainable Development Goals (UN SDGs) at a Glance. *Big Data Cogn. Comput.* **2021**, *5*, 28. [CrossRef]

42.   Asaithambi, S.P.R.; Venkatraman, S.; Venkatraman, R. Big Data and Personalisation for Non-Intrusive Smart Home Automation. *Big Data Cogn. Comput.* **2021**, *5*, 6. [CrossRef]

43.   Choi, J.-y.; Cho, M.; Kim, J.-S. Employing Vertical Elasticity for Efficient Big Data Processing in Container-Based Cloud Environments. *Appl. Sci.* **2021**, *11*, 6200. [CrossRef]

44.   Shah, S.A.R.; Waqas, A.; Kim, M.-H.; Kim, T.-H.; Yoon, H.; Noh, S.-Y. Benchmarking and Performance Evaluations on Various Configurations of Virtual Machine and Containers for Cloud-Based Scientific Workloads. *Appl. Sci.* **2021**, *11*, 993. [CrossRef]

45.   Azeroual, O.; Fabre, R. Processing Big Data with Apache Hadoop in the Current Challenging Era of COVID-19. *Big Data Cogn. Comput.* **2021**, *5*, 12. [CrossRef]

46.   Burgin, M.; Mikkilineni, R. From data Processing to Knowledge Processing: Working with Operational Schemas by Autopoietic Machines. *Big Data Cogn. Comput.* **2021**, *5*, 13. [CrossRef]

47.   Almohsen, K.; Al-Jobori, H. Recommender Systems in Light of Big Data. *Int. J. Electr. Comput. Eng.* **2015**, *5*, 1553–1563. [CrossRef]

48.   Fayyaz, Z.; Ebrahimian, M.; Nawara, D.; Ibrahim, A.; Kashef, R. Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Appl. Sci.* **2020**, *10*, 7748. [CrossRef]

49.   Verma, J.P.; Patel, B.; Patel, A. Big Data Analysis: Recommendation System with Hadoop Framework. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence & Communication Technology, Ghaziabad, India, 13–14 February 2015; pp. 92–97. [CrossRef]

50.   Kaufmann, M. Big Data Management Canvas: A Reference Model for Value Creation from Data. *Big Data Cogn. Comput.* **2019**, *3*, 19. [CrossRef]

51.   Ajah, I.A.; Nweke, H.F. Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data Cogn. Comput.* **2019**, *3*, 32. [CrossRef]

52.   Brynjolfsson, E.; Hitt, L.M.; Kim, H.H. Strength in Numbers: How Does Data-Driven Decision Making Affect Firm Performance? 2011. Available online: http://ssrn.com/abstract=1819486 (accessed on 10 August 2021).

53.   International Data Corporation (IDC). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. 2014. Available online: http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm (accessed on 10 August 2021).

54.   Hilbert, M.; López, P. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* **2011**, *332*, 60–65. [CrossRef]

55.   Russom, P. *Managing Big Data*; TDWI–The Data Warehousing Institute: Los Angeles, CA, USA, 2013.

56.   Brynjolfsson, E.; McAfee, A. Big Data: The Management Revolution. *Harv. Bus. Rev. Press* **2012**, *90*, 60–68.

57.   Westerman, G.; Calméjane, C.; Bonnet, D.; Ferraris, P.; McAfee, A. *Digital Transformation: A Roadmap for Billion-Dollar Organizations*; MIT Center for Digital Business and Capgemini Consulting: Paris, France; Cambridge, MA, USA, 2011; pp. 1–68.

58.   Boueé, C.; Schaible, S. *Die Digitale Transformation der Industrie*; Studie:Roland Berger und BDI: Munich, Germany, 2015.

59.   Grimmer, J.; Steward, B.M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Anal.* **2013**, *21*, 1–31. [CrossRef]

60.   Roberts, C.W. A Conceptual Framework for Quantitative Text Analysis. *Qual. Quant.* **2000**, *34*, 259–274. [CrossRef]

61.   Marshall, C.; Rossman, G.B. *Designing Qualitative Research*, 5th ed.; Sage: Thousand Oaks, CA, USA, 2015.

62.   Berg, B.L. *Qualitative Research Methods for the Social Sciences*; Bacon & Allyn: Boston, MA, USA, 1995.

63.   Randolph, J.J. A guide to writing the dissertation literature review. *Pract. Assess. Res. Eval.* **2009**, *14*, 1–13.

64.   vom Brocke, J.; Simons, A.; Niehaves, B.; Reimer, K.; Plattfaut, R.; Cleven, A. Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In Proceedings of the 17th European Conference on Information Systems, ECIS 2009, Verona, Italy, 8–10 June 2009; p. 161.

65.   Fisher, D.; DeLine, R.; Czerwinski, M.; Drucker, S. Interactions with big data analytics. *Interact. ACM* **2012**, *19*, 50–59. [CrossRef]

66.   Bizer, C.; Boncz, P.; Brodie, M.L.; Erling, O. The Meaningful of Big Data: Four perspectives—Four challenges. *SIGMOD Record* **2011**, *40*, 56–60. [CrossRef]

67.   Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Byers, A.H. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*; McKinsey Global Institute: New York, NY, USA, 2011.

68. Zhu, H.; Xu, Z.; Huang, Y. Research on the security technology of big data information. In Proceedings of the International Conference on Information Technology and Management Innovation, Singapore, 3–6 February 2015; pp. 1041–1044.
69. Hongjun, Z.; Wenning, H.; Dengchao, H.; Yuxing, M. Survey of Research on Information Security in Big Data. In Proceedings of the Congresso da Sociedada Brasileira de Computacao, Brasilia, Brasil, 22–25 July 2014; pp. 1–6.
70. Zan, M.; Yanfei, L. Research of Big Data based on the views of technology and application. *Am. J. Ind. Bus. Manag.* **2015**, *5*, 192–197.
71. Trifu, M.; Ivan, M. Big Data: Present and future. *Database Syst. J.* **2014**, *5*, 32–41.
72. Sathi, A. *Big Data Analytics. Distributive Technologies for Changing the Game*; Mc Press: Boise, ID, USA, 2012.
73. Laney, D. *Big Data Means Big Business*; Gartner Inc.: Stamford, CT, USA, 2013.
74. Deloitte. *Industry 4.0, Challenges and Solutions for the Digital Transformation and Use of Exponential Technologies*; Deloitte AG: London, UK, 2015; pp. 1–32. Available online: https://www.pac.gr/bcm/uploads/industry-4-0-deloitte-study.pdf (accessed on 12 September 2021).
75. Khare, A.; Khare, K.; Baber, W. Why Japan's Digital Transformation Is Inevitable. In *Transforming Japanese Business: Rising to the Digital Challeng*; Springer: Singapore, 2020. [CrossRef]
76. EU. ENTR/E4-Fuelling Digital Entrepreneurship in Europe. Background Paper, European Commission 2018, EU Commission Strategic Policy Forum on Digital Entrepreneurship. Available online: http://ec.europa.eu/DocsRoom/documents/5313/attachments/1/translations (accessed on 10 January 2019).