


# A Patent Analysis on Big Data Projects

Gustavo Grander, Universidade Nove de Julho, Brazil

Luciano Ferreira da Silva, Universidade Nove de Julho, Brazil

 <https://orcid.org/0000-0001-6482-8729>

Ernesto D. R. Santibanez Gonzalez, University of Talca, Chile

## ABSTRACT

Studies concerning big data patents have been published; however, research investigating big data projects is scarce. Therefore, the objective of this study was to conduct an exploratory analysis of a patent database to collect information about the characteristics of registered patents related to big data projects. The authors searched for patents related to big data projects in the Espacenet database on January 10, 2021 and identified 109 records. The textual analysis detected three word classes interpreted as (1) a direction to cloud computing, (2) optimization of solutions, and (3) storage and data sharing structures. The results also revealed emerging technologies such as blockchain and the internet of things, which are utilized in big data project solutions. This observation demonstrates the importance that has been given to solutions that facilitate decision-making in an increasingly data-driven context. As a contribution, they understand that this study endorses a group of researchers that has been dedicated to academic research on patent documents.

## KEYWORDS

Big Data, Blockchain, Cloud, Intellectual Property, IoT, Patent, Project, Technology

## 1. INTRODUCTION

Big Data is characterized as a large data set, which is challenging to store, process, analyze, and understand using traditional database processing tools (S. Huang & Chaovaitwongse, 2015). The advancement in the use of Big Data has driven a generation of technologies and architectures designed to extract economic value through analysis (Gantz & Reinsel, 2011). Thus, decisions oriented to data analysis stand out compared to decisions based on intuition (McAfee & Brynjolfsson, 2012). With the increase in the generation of unstructured data, the analysis and interpretation through structured tables with rows and columns becomes increasingly complex.

A Big Data project, in its turn, can be defined as a data-intensive project that presents large-scale problems with restrictions of volume, variety, speed (Becker, 2017), and still veracity of the data (Barham & Daim, 2018). Therefore, the culture of data-driven decision-making is essential to the success of a Big Data project (Dutta & Bose, 2015). Big data projects, using cutting-edge analysis involving artificial intelligence and machine learning, condition computers to identify what these data represent by identifying patterns more assertively.

Decision-making based on data depends on the successful implementation of Big Data projects (Dutta & Bose, 2015). However, the excessive focus on data and the neglect of adequate decision-

DOI: 10.4018/IJBAN.288516

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

making has been evidenced as an aspect that reflects the failure of Big Data projects (Chiheb, Boumahdi, & Bouarfa, 2019). Big Data project characteristics reflect potential highly uncertain inputs (for example, the relevance of data) and also potentially highly uncertain results (for example, information may be inaccurately derived from data analysis) (J. Saltz, Shamshurin, & Connors, 2017). Big Data projects have also provided a good discussion regarding success factors. In a broad perspective, there is a convergence between studies that investigated success factors in big data projects and that encompass five predominant characteristics: Governance, Objectives, Processes, Data, Team, and Tools (Gao, Koronios, & Selle, 2015; J. S. Saltz & Shamshurin, 2016).

Despite the evolution made in the Big Data area in recent years, practitioners reinforce the need for research related to guides, models, or even methodologies (Lara, De Sojo, Aljawarneh, Schumaker, & Al-Shargabi, 2020). As a result, opportunities arise for project management professionals to contribute by identifying appropriate measures and types of data, helping to interpret the data, and placing them in an evaluation context (Olsson & Bull-Berg, 2015). In addition to identifying technological trends in specific fields (Abbas, Bilal, Zhang, & Khan, 2015), patent analysis can identify recurring problems in a specific area, as patent registrations are intended to solve a practical problem.

Previous patent analysis studies of Big Data technology have been performed. Examples include abstract analysis of Chinese Big Data patents (Z. Yang et al., 2017), hot classification fields of Big Data technology (Gui, Liu, Bai, & Zhang, 2017), methods technology assessment using quantitative patent analysis for technology transfer in Big Data marketing (Jun, Park, & Jang, 2015), business interests and activities around Big Data (Y. Huang et al., 2016) and an analysis of the patenting activities of global jurisdictions in the Big Data field (Saheb & Saheb, 2020). While these studies analyze Big Data patents, they do not directly address the specific area of Big Data projects.

Singh (2019) stated that Big Data projects involve research in purely technical areas, such as ubiquitous information and integration in Big Data ecosystems. Thus, given the relevance of Big Data projects and the scarcity of studies characterizing the patents generated in this area, we took this as an opportunity for new research. Herein, an exploratory analysis of the Espacenet patent database was performed on January 10, 2021. to identify and subsequently describe registered Big Data projects patents. based on an initial sample of 109 records.

Our results advance the discussion on Big Data projects since this study fills a gap in intellectual property knowledge. This study also endorses the researchers who promote patent research to search for information on solving technical problems without being limited to scientific articles. Finally, based on this study's main findings, we suggest extending this analysis from a Big Data project management perspective in the future.

In the following sections, we present the Theoretical background (Section 2), Methodological procedures (Section 3), Results and Discussion (Section 4), and Conclusions (Section 5) of this study.

## **2. THEORETICAL BACKGROUND**

Previous studies reported an approach for organizations to create Big Data projects, encompassing phases of project planning, implementation, and post-implementation (Mousannif, Sabah, Douiji, & Sayad, 2014, 2016). Moreover, another study proposed implementing Big Data projects based on the strategic basis, data analysis, and implementation phases (Dutta & Bose, 2015). These two proposals, which aim to establish effective Big Data solutions, initially discuss the strategic aspect of the project and the attendance of business problems, reinforcing the holistic characteristic of all Big Data projects.

Interestingly, the strategic alignment of Big Data projects with the organizational vision has been considered a success factor in Big Data projects (Gao et al., 2015; J. S. Saltz & Shamshurin, 2016). Indeed, the proper use of data visualization technologies is crucial for successful data interpretation and must be considered by the decision-makers at the strategic level, such as data analysts and executives (Moore, 2017).

Big data and cloud computing are among the technological revolutions of the moment and have led to significant transformations in information technology and impacts on scientific research (Boutkhoul, Hanine, Agouti, & Tikniouine, 2016). Cloud computing provides vital support to address the challenges related to shared computing resources such as computing, storage, networking, and analytical software. Notably, the application of these resources has led to impressive advances in Big Data (C. Yang, Huang, Li, Liu, & Hu, 2017).

Some emerging technologies rely on Big Data. For example, the Internet of Things (IoT) is a network of heterogeneous devices with sensors with different functionalities connected to a private or public network (Miorandi, Sicari, De Pellegrini, & Chlamtac, 2012). Its ability to react to the surrounding environment depends on Big Data. Additionally, Blockchain, a distributed public accounting technology initially intended for cryptocurrencies, utilizes Big Data technologies. It has been reported that systems that combine blockchain technology and the IoT benefit from lower operating costs, decentralized resource management and robustness against threats and attacks. Therefore, the convergence of the IoT and Blockchain technologies could overcome significant challenges currently faced by the IoT (Ferrag et al., 2018).

In the information age, data has become an essential strategic resource. In this sense, Big Data, especially in the IoT era, has become a practical tool for enterprise and government decision-making, specifically impacting economic and social life (Wang, 2020).

Data mining is the computational process of discovering patterns in large data sets. It involves methods at the intersection of Artificial Intelligence (AI), machine learning, statistics, and database systems (Majumdar, Naraseeyappa, & Ankalaki, 2017). According to Fayyad, Piatetsky-Shapiro, and Smyth (1996), data mining applies specific algorithms or techniques to extract knowledge patterns from data sets. The patterns discovered should be significant insofar as they lead to some advantage in decision-making (Ahmed, Aziz, Tezel, & Riaz, 2018).

Decision-making models for Big Data projects have been proposed to reduce problems of limited rationality in the context of complex decisions. Chiheb et al. (2019) presented a conceptual model called BD-Da that describes the levels to be considered to develop a Big Data project to solve the often neglected decision-making problem. The model helps companies make faster and more intelligent decisions and gain a competitive advantage.

Elgendy and Elragal (2016) developed a B-DAD structure that integrates Big Data tools, architecture, and analysis. The structure was developed and tested to identify promotional products that should be offered for sale. It was also shown to make decision-making more assertive through advanced Big Data analytical techniques that reveal hidden insights. Later, Poletto, De Carvalho, and Costa (Poletto, De Carvalho, & Costa, 2017) presented an integrated model of Big Data, Business Intelligence, and Decision Support System to apply during the decision-making phases. The authors state that organizations are essentially composed of processes, people, and technology; however, technology alone is useless without one of the first two.

More specific solutions within Big Data projects have also been developed. Examples include a decision-making approach based on the diffuse AHP-TOPSIS methodology to select an appropriate cloud solution to manage Big Data projects (Boutkhoul, Hanine, Agouti, & Tikniouine, 2017). Additionally, a project management and execution support platform with an emphasis on analytics and reporting was reported to facilitate successful project decision-making (Pondel & Pondel, 2016). It is also evident that there is a need for greater criticality for the data storage and analysis steps, especially concerning the high risk of collecting low-quality data and the steep learning curve of the initial phase of Big Data projects (Lara et al., 2020).

Large data sets originating from heterogeneous to distributed sources likely have different data formats and models. They may not be in the same data format or model of a target Data Warehouse or contain missing, low quality, redundant or data. Thus, Data Warehouse architecture software, known as Extraction-Transformation-Loading (ETL), is employed (Ali & Wrembel, 2017). The emergence

of Big Data has caused more and more companies to modify their data management strategies to optimize performance (Jianmin, Wenbin, Tongrang, Shilong, & Hongwei, 2020).

After highlighting the relevant discussion based on the studies cited in the context of Big Data projects, in the next section, we describe the methodological procedures of our research and provide details about data collection and analysis.

### 3. METHODOLOGICAL PROCEDURES

According to Clarke (2018), depending on the questions you are being asked (i.e., state of the art, prior art/novelty/patentability, freedom to operate, opposition, and (in)validity), there are different types of patent searches. Herein, we sought to collect relevant documents published in a particular technical field and extract data and descriptive information.

To achieve our goal of carrying out an exploratory analysis of the characteristics of registered Big Data projects patents, we searched the Espacenet database on January 10, 2021, using the following search string: “Big data project\*”. The search retrieved 109 records, and after a structured analysis of the documents, six (06) patents were eliminated from the sample because they did not have a Big Data project scope or cited the search term in a decontextualized way. We also identified two cases of repetitions, in which the patents were registered in two offices. It should be pointed out that for the descriptive analysis, we did not consider the six patents outside the scope of the research. For the textual lexical analysis, we did not consider the six patents outside the scope or the two repeated patents to not bias the quality of the results (Table 1).

Table 1. Patents eliminated and repeated from the sample

	Patent number	Patent name
Eliminated patents	CN108446206A	AR technology-based application availability display system
	US2020324135A1	Device and method for treating eight cancers with ultraviolet radiation
	CN107177625A	Artificial carrier system with site-specific mutagenesis and site-specific mutagenesis method
	CN107671020A	Automatic cleaning brush of seabed observation instrument equipment and work principle of automatic cleaning brush
	CN104484582A	Method and system for automatically analyzing bio information items through modular selection
	WO2019100012A1	Data collection & analytics based on detection of biological cells or biological substances
Repeated patents	US9384264B1	Analytic systems, methods, and computer-readable media for structured, semi-structured, and unstructured documents
	EP3179387A1	
	WO2020014399A1	Decentralized Cybersecure Privacy Network For Cloud Communication, Computing And Global e-Commerce
	US2019386969A1	

For data analysis, we used Iramuteq version 0.7, free and open-source software that enables the textual analysis of a corpus prepared by the researchers. We organized patent summaries according to the available information in the documents. After processing the corpus, 230 text segments and 7.890 occurrences were identified. Only adjectives, verbs, and common names were used as analysis keys.

An analysis of the textual corpus, consisting of 230 text segments and 7,890 occurrences from 101 textual patents, identified 567 hápax or words that appear only once in the entire textual corpus. A summary of the corpus is presented in Table 2.

**Table 2. Summary of corpus analysis**

Summary of corpus analysis	Total
Number of texts	101
Number of text segments	230
Occurrences	7,890
Average occurrence per text	78.90
Number of forms	1,269
Number of hápax	567*

\*44.68% of form, 7.19% of occurrences.

Next, we used the Descending Hierarchical Classification Method (DHCM) to obtain text segment classes. At the same time, DHCM detects text segment vocabulary that is similar and different in the classes (Camargo, 2005). We also performed a Factorial Correspondence Analysis (FCA), a feature of Iramuteq that produces graphical representations in a Cartesian plane. FCA helps visualize the proximity and distances of words from the crossing of the textual corpus vocabulary. The procedures performed in this analysis include the calculation of frequencies, chi-square correlation values for each word in the corpus, and the execution of the FCA in a contingency table that crosses the active forms and the variables (Salviati, 2017).

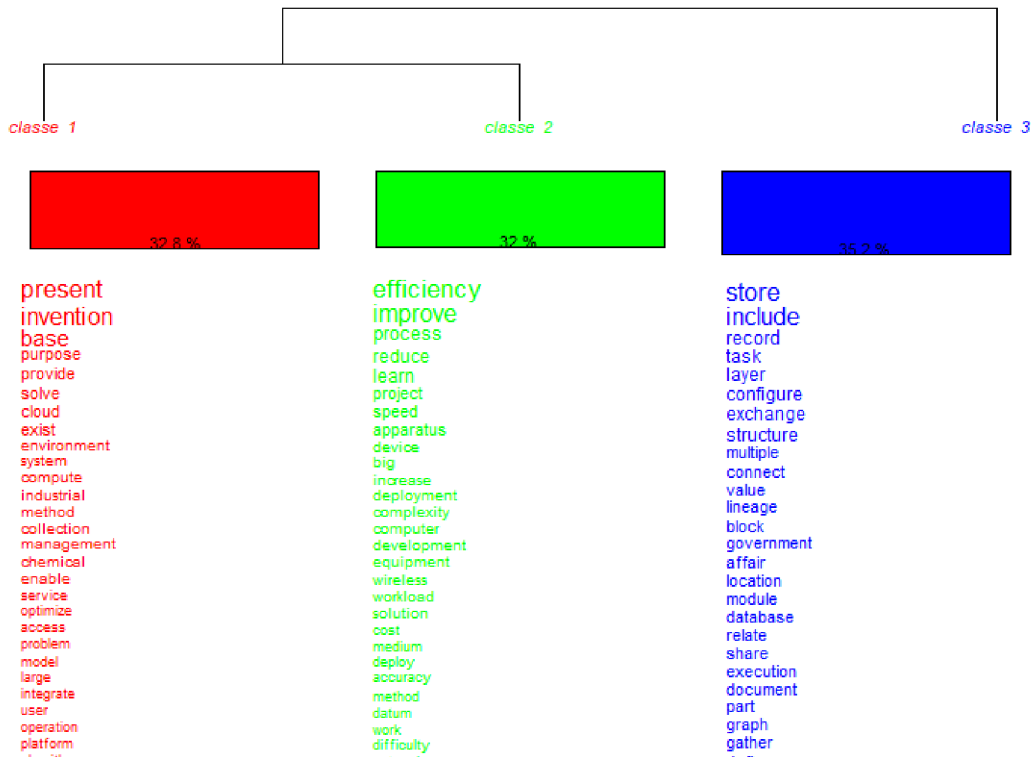
## 4. RESULTS AND DISCUSSIONS

As shown in Table 3, China displayed an expressive number of registrations in China with 81 patents, corresponding to 78.64% of the sample. China registered one patent in 2015 and has exhibited year-to-year growth until 2020. The United States of America (USA) is the country with the second most registrations with ten records, and the World Intellectual Property Organization (WIPO) is third with six records. Moreover, the Republic of Korea (ROK) had three records, the European Patent Office (EPO) had two, and Japan had one. The large number of registrations and steady increase in productivity over the years highlight China's interest in the topic.

**Table 3. Intensity map of the number of records in countries/regions related to big data projects**

Country/Region	2015	2016	2017	2018	2019	2020	Total
China	1	2	6	13	17	42	81
USA	2	1	1		3	3	10
WIPO		1		1	1	3	6
ROK				1		2	3
EPO			1		1		2
Japan						1	1
Total	3	4	8	15	22	51	103

Figure 1. Dendrogram of DHCM



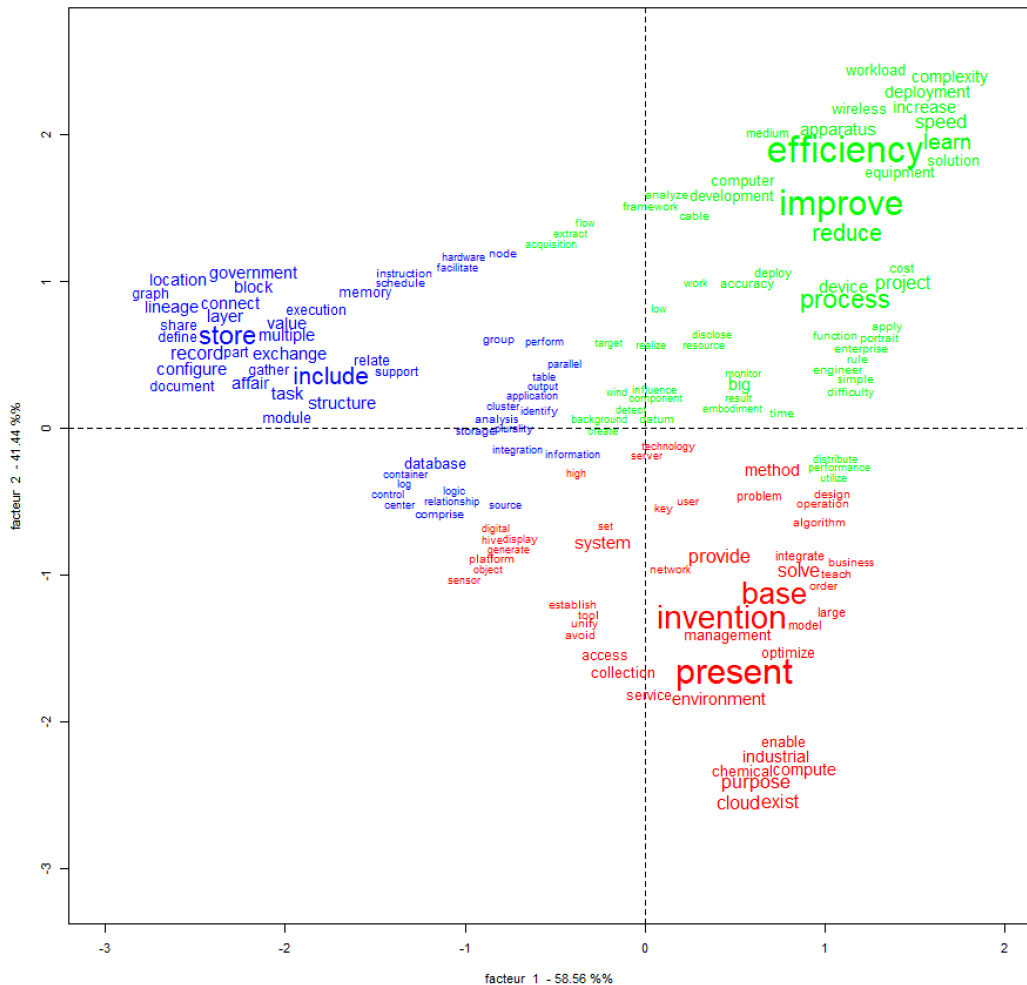
When studying Big Data patents in China, Z. Yang et al. (2017) showed that from 1980–2016 the increase in Big Data patent registrations in China occurred after 2005, and the claimants' patent applications were mainly from universities and companies. Teixeira and Queirós (2016) highlight the 'Intermediate Intellectual Property Service', which since 2008 has been a strategy for technological development to make China a country with a high level of intellectual property creation. Indeed, our intensity map illustrates the predominance of registrations in China, confirming the success of this strategy.

When generating the dendrogram (Fig. 1), which is a tree diagram that displays the groups formed by clustering at their similarity levels, three classes were identified and analyzed according to the relationships between the words. Class 1 represents 32.8% of the corpus and has words like "system", "method", "model" and "management" that are related to the corpus as procedures or forms of organization. Class 1 also contains words like "cloud", "compute", "platform" and "service" that can indicate a direction for the use of cloud computing services. Class 2 represents 32% of the corpus and consists of words such as "efficiency", "improve", "process", "reduce", "speed" and "solution" suggesting the corpus is directed towards the optimization of solutions. Additionally, Class 2 also contains words like "apparatus", "device", "equipment" and "wireless", referring to objects that facilitate solution optimization. Lastly, Class 3 represents 35.2% of the corpus and is characterized by words like "store", "database", "block", "record", "structure", "module" and "share", which are indicative of data storage and sharing structures.

It should be pointed out that since Classes 1 and 2 have equivalence in terms of word means compared to Class 3; thus, they could have been combined into one class.

The FCA (fig. 2) made it possible to visualize the textual corpus distributed in the Cartesian plane. As shown in Figure 2, the three distinct classes do not overlap. The textual corpus was formed by the

Figure 2. FCA of the sample



words in the patent summary, and each word had to be repeated at least ten times to be considered in the FCA. The interpretation of word placement is subjective, and it is up to researchers to understand why the words have this distribution. Thus, after the software generated the FCA, we carried out a deep analysis directed at all patents by their respective words to identify relationships between them and form a common corpus that justifies each word group.

When interpreting the distances between words and distributions along the plane, the x-axis appears to refer to data collection and storage technologies (left) and processing and performance techniques (right). In other words, patents related to Big Data design applications. On the other hand, the y-axis indicates a tangibility relationship such that the lower and upper parts of the plane represent intangible and tangible solutions, respectively.

Due to continuous development, IoT technology and Big Data have become analytical tools widely applied in many technical fields (Wang, 2020). The rapid evolution of decision support systems has made experts realize that putting Big Data on the cloud is a real challenge that businesses must

consider. In this sense, the selection of cloud solutions is recognized as an essential research issue for Big Data projects (Boutkhoul et al., 2016).

A significant challenge faced by this sector is that Big Data generated from the various IoT sensors contains rich spatiotemporal information (C. Yang et al., 2017). We identified a Big Data IoT project in the cloud (Patent number CN112001023A) which provides a method for analyzing the degree of safety from pollution by measuring reflected light from a glass curtain wall. According to the authors, the invention helps to continuously monitor the light pollution phenomenon on the glass curtain wall. Furthermore, utilizing Big Data, the patent objectively analyzes and visualizes the security of the building's light environment for protecting the eyesight and improving the productivity of the employees.

It has been reported that the advancement of IoT is expected to provide easy public access to valuable results from Big Data and cloud computing (C. Yang et al., 2017). However, access to the best information needs to be provided to the public for better decision-making (Abbas et al., 2015). Herein, we found Patent number CN111338302A that refers to the technical field of Smart factories, and more specifically, to a chemical modeling processing system based on industrial Big Data and IoT. For inventors, the invention aims to integrate and implement AI and Big Data technology in an independent and controllable way into traditional production lines, helping production companies achieve synergistic data, production, and control flows, improve production efficiency, reduce production costs. In other words, this technology could contribute to the development of intelligent, independent and controllable manufacturing.

It is predicted that Blockchain technology will soon overcome significant challenges faced in this area of research and development. In this context, a method for extracting and managing project results generated by specialists, executed by a server and stored in a Blockchain has been developed (Patent number WO2020175753A1). Moreover, we also identified a Blockchain consensus algorithm and platform, Blockchain encryption technology, and evolutionary network developed to reduce cryptocurrency transaction costs and time (Patent number KR20200045089A). Alkhamisi and Alboraei (2019) state that with the rise of Blockchain technology, recent research has focused on adopting Blockchain to address security issues associated with IoT. Furthermore, Ferrag et al. (2018) classified threat models considered by Blockchain protocols on IoT networks into five main categories: identity-based attacks, manipulation-based attacks, cryptanalytic attacks, reputation-based attacks, and service-based attacks.

In the sample, we also detected patents related to a remote guidance control system for corporate services (patent number CN107748764A). This invention makes querying a company's service demand information more efficient and convenient by efficiently combining the client, the platform cloud data, and third-party service agencies. Despite the numerous advantages, it is challenging to find recommendation systems with a specific formula that directly provide optimized parameters. This setback is mainly attributed to the fact that more and more parameters are generated as the machine learning models become more and more complex. Notably, it has been proposed that a minimum gradient descent algorithm based on the Gene, Behavior, and Phenotypic (GBP) dimension could mitigate this problem (patent number CN111581524A). Recommendation systems collect historical user classifications, interactions, portraits, social networks, and context and analyze historical interests to recommend relevant information for users to meet needs and recommendation algorithms, which are the core of recommendation systems.

Data mining plays a crucial role in analyzing large data sets. Patent number US10282603B2 refers to a method implemented by a computer to analyze technical documents concerning a corpus. According to the inventors, several algorithms can perform data searches; however, information retrieval is often an inaccurate process with the rise of Big Data projects. Therefore, improvements in information retrieval are necessary. We also found a series of patents related to performance optimization. For example, patent number CN103823881A describes a method and device for optimizing the performance of a distributed database. Additionally, patent number CN106130801A



provides a method for updating and monitoring Big Data, which solves the problems associated with components during traditional distributed data collection.

When an ETL process is implemented as a workflow, where tasks process data and are connected by data flows, ETL tools must be developed to execute these workflows (Ali & Wrembel, 2017). Patent number CN107145576A is an ETL tool that addresses a Big Data ETL scheduling system by supporting visualization and process flow, eliminating complicated background operations, improving ETL development speed and efficiency, and reducing implementation costs. It has been reported that the ETL architecture strategy processes the data uniformly (Jianmin et al., 2020). Moreover, patent CN107330028A discloses an Apache NiFi application extension method for importing source data into a database. According to the method, different access tools are not required when multiple data sources have access to the database. Consequently, the data access process is unified. An access plug-in can also be extended automatically.

The culture of data-driven decision-making has promoted structural changes in processes, people, and technology (Elgendy & Elragal, 2016). The value built from the use of Big Data in the decision-making process lies in the ability of decision-makers to allocate resources and make adjustments to the process more assertively. Indeed, some patents identified in our sample potentially improve decision-making. However, as Boutkhoul et al. (2017) pointed out, selecting the most appropriate solution for Big Data projects is complex and requires a comprehensive evaluation process.

Patent CN106096733A presents an analytical model that evaluates Big Data for water conservation. This method applies an analytical hierarchy process and a fuzzy mathematical theory to Big Data, comprehensively considering various schemes of water service. Once the schemes are classified, the best decision-making scheme is selected. Another patent (CN11144446A) refers to a method and system for collecting engineering cost data based on Big Data, computer equipment, and a computer-readable storage medium. According to the inventors, the method has the advantage that different decision-makers can master the cost data for each stage in each engineering field conveniently. The total engineering cost is obtained as a whole, and, finally, a reasonable decision is made.

We also highlight patent number CN111652090A, which belongs to the mechanical vision field. This invention discloses a digital management method for planting hemp in China. China's hemp planting companies do not monitor the main control points, and manual operation often generates some hysteresis information. Consequently, the timing of decision-making and deviation correction actions are negatively affected. This invention uses a video library database of plant growth images and monitors the images to build a model library and knowledge base. The system can perform batch image processing and recognition by configuring parameters to assist decision-making and improve productivity.

Big Data in modern businesses allows companies to make faster and wiser decisions and achieve competitive advantages. However, many Big Data projects provide disappointing results that do not meet the needs of the decision-maker. These poor results are because they neglect the decision of these projects, which is an aspect of the decision-making process. (Chiheb et al., 2019). It has been stated that with the growing demand for using Big Data to take advantage of their opportunities, organizations are looking for clear and simple solutions and guidelines for Big Data management (Elgendy & Elragal, 2016). As highlighted by Poletto, Carvalho, and Costa (2017), the use of Big Data will only be successful if there is a data integration strategy for the generation of relevant information and knowledge management.

Finally, we highlight applicants with more than one record in our sample. Altogether nine applicants fit this requirement, including Limited Companies with 11 records, followed by universities with two, a research institute with one, and an individual with one (Table 4).

Patent data presents valuable insights for inventors, engineers, companies, and decision-makers (Akers, 2003). It is also considered an indicator of innovation and reflects technological and scientific changes (Archibugi, 1992). In this sense, the mapping of inventions and, consequently, the applicants can reveal strong indicators of technological trends and, in some cases, mastery of a technique or

**Table 4. Applicants with more than one registered patent**

Applicants	Patent number	Patent name
Chongqing Socialcredits Big Data Tech Co Ltd	CN109408712A	A method for constructing a multi-dimensional information portrait of a travel agency user
	CN109345348A	Recommendation method of multi-dimensional information portrait based on travel agency users
Electric Power Res Inst	CN109190381A	A detection method for Hadoop security vulnerability
	CN109961376A	Distributed energy storage equipment management and control system and method
H3C Big Data Tech Co Ltd	CN108062399A	Data processing method and device
	CN109344163A	Data verification method, apparatus, and computer-readable medium
Lenovo Beijing Co Ltd	CN108804233A	Memory space recycling method, device, electronic equipment and storage medium
	CN106293929A	Data processing method and first electronic equipment
Shenzhen OneConnect Intelligent Technology Co Ltd	CN111984685A	Data tilt detection method, device, computer equipment and readable storage medium
	CN111984630A	Log association method, device and computer equipment
	CN111158642A	Data construction method and device, computer equipment and storage medium
South China University of Technology	CN111444236A	Mobile terminal user portrait construction method and system based on Big Data
	CN107817787A	Intelligent production line manipulator fault diagnosis method based on machine learning
Xiamen University	CN207423929U	Novel air quality intelligence prediction system
	CN111310868A	Water-writing handwritten character recognition method based on convolutional neural network
Wang Yujue	CN109650186A	Communication cable taking-up and paying-off device for communication Big Data project
	CN109650187A	Communication cable taking-up and paying-off device for communication Big Data project
Xinhua Sanda Data Tech Co Ltd	CN107548151A	Method and device for marking wireless access point (AP)
	CN107547519A	Wireless Big Data deployment method and device

technology. Therefore, both academia and industry must recognize the emerging patterns of big data technologies, which are of primary importance in the growth of data-driven enterprises (Saheb & Saheb, 2020).

## 5. CONCLUSION

When carrying out an exploratory analysis with patent data to understand the characteristics of patent registration related to Big Data projects, we observed a significant number of patents registered in China. This result indicates substantial importance this country gives to the subject. Notably, the dendrogram analysis and FCA identified three classes of words interpreted as a direction to cloud computing, optimization of solutions, and storage and data sharing structures.

The main limitation of our study was that despite the Espacenet database being one of the primary databases, with more than 120 million patent documents, incremental research with other databases could complement the study sample and possibly produce results complementary to those presented herein.

The present study also found that emerging technologies such as Blockchain and IoT are being used in Big Data project solutions. Moreover, we observed the importance given to solutions that help decision-making in an increasingly data-driven context. Indeed, this is perhaps the most significant practical contribution of this study.

As an academic contribution, this study endorses a body of researchers that have dedicated their academic research to patent documents, making patent studies in the academic environment more noticeable.

Since Big Data research, and more specifically, Big Data projects, are areas of incipient knowledge, there are many opportunities for new research on the topic. This observation is the main implication of this paper. We suggest that future research extend this discussion to Big Data project management through a theoretical lens or even applied research for empirical validation of this study's findings. Based on the identified patent clusters, it is plausible that empirical research could explain the Big Data project phenomenon.

## **ACKNOWLEDGMENT**

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

## REFERENCES

- Ahmed, V., Aziz, Z., Tezel, A., & Riaz, Z. (2018). Challenges and drivers for data mining in the AEC sector. *Engineering, Construction, and Architectural Management*, 25(11), 1436–1453. doi:10.1108/ECAM-01-2018-0035
- Akers, L. (2003). The future of patent information - A user with a view. *World Patent Information*, 25(4), 303–312. doi:10.1016/S0172-2190(03)00086-3
- Ali, S. M. F., & Wrembel, R. (2017). From conceptual design to performance optimization of ETL workflows: Current state of research and open problems. *The VLDB Journal*, 26(6), 777–801. doi:10.1007/s00778-017-0477-2
- Alkhamisi, A. O., & Alboraei, F. (2019). Privacy-aware Decentralized and Scalable Access Control Management for IoT Environment. *Journal of King Abdulaziz University Computing and Information Technology Sciences*, 8(1), 71–84. doi:10.4197/Comp.8-1.7
- Archibugi, D. (1992). Patenting as an indicator of technological innovation: A review. *Science & Public Policy*, 19(6), 357–368. doi:10.1093/spp/19.6.357
- Barham, H., & Daim, T. (2018). Identifying critical issues in smart city big data project implementation. *Proceedings of the 1st ACM/EIGSCC Symposium on Smart Cities and Communities, SCC 2018*. doi:10.1145/3236461.3241967
- Becker, D. K. (2017). Predicting outcomes for big data projects: Big Data Project Dynamics (BDPD): Research in progress. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, 2320–2330. doi:10.1109/BigData.2017.8258186
- Boutkhoul, O., Hanine, M., Agouti, T., & Tikniouine, A. (2016). Selection problem of Cloud solution for big data accessing: Fuzzy AHP-PROMETHEE as a proposed methodology. *Journal of Digital Information Management*, 14(6), 368–382.
- Boutkhoul, O., Hanine, M., Agouti, T., & Tikniouine, A. (2017). A decision-making approach based on fuzzy AHP-TOPSIS methodology for selecting the appropriate cloud solution to manage big data projects. *International Journal of Systems Assurance Engineering and Management*, 8(S2), 1237–1253. doi:10.1007/s13198-017-0592-x
- Camargo, B. V. (2005). *Alceste: Um programa informático de análise quantitativa de dados textuais*. E. Universitária, Ed.
- Chiheb, F., Boumahdi, F., & Bouarfa, H. (2019). A conceptual model for describing the integration of decision aspect into big data. *International Journal of Information System Modeling and Design*, 10(4), 1–23. doi:10.4018/IJISMD.2019100101
- Clarke, N. S. (2018). The basics of patent searching. *World Patent Information*, 54, S4–S10. doi:10.1016/j.wpi.2017.02.006
- Dutta, D., & Bose, I. (2015). Managing a big data project: The case of Ramco cements limited. *International Journal of Production Economics*, 165, 293–306. doi:10.1016/j.ijpe.2014.12.032
- Elgendy, N., & Elragal, A. (2016). Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science*, 100, 1071–1084. doi:10.1016/j.procs.2016.09.251
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–53.
- Ferrag, M. A., Derdour, M., Mukherjee, M., Derhab, A., Maglaras, L., & Janicke, H. (2018). Blockchain technologies for the internet of things: Research issues and challenges. *ArXiv*, 6(2), 2188–2204.
- Gantz, J., & Reinsel, D. (2011). Extracting Value from Chaos State of the Universe: An Executive Summary. *IDC IView*, (June), 1–12. Retrieved from <http://idcdocserv.com/1142>
- Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in Big Data analytics projects. *2015 Americas Conference on Information Systems, AMCIS 2015*, 1–14.
- Gui, B., Liu, Y., Bai, X., & Zhang, J. (2017). Longitudinal patent analysis for big data technology. In *PICMET 2017 - Portland International Conference on Management of Engineering and Technology: Technology Management for the Interconnected World, Proceedings* (pp. 1–8). doi:10.23919/PICMET.2017.8125461

- Huang, S., & Chaovalitwongse, W. A. (2015). Computational Optimization and Statistical Methods for Big Data Analytics: Applications in Neuroimaging. *The Operations Research Revolution*, (November), 71–88. doi:10.1287/educ.2015.0135
- Huang, Y., Youtie, J., Porter, A. L., Robinson, D. K. R., Cunningham, S. W., & Zhu, D. (2016). Big data and business: Tech mining to capture business interests and activities around big data. In *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016* (pp. 145–150). doi:10.1109/BDCloud-SocialCom-SustainCom.2016.32
- Jianmin, W., Wenbin, Z., Tongrang, F., Shilong, Y., & Hongwei, L. (2020). An improved join-free snowflake schema for ETL and OLAP of data warehouse. *Concurrency and Computation*, 32(23), 1–11. doi:10.1002/cpe.5519
- Jun, S., Park, S., & Jang, D. (2015). A Technology Valuation Model Using Quantitative Patent Analysis: A Case Study of Technology Transfer in Big Data Marketing. *Emerging Markets Finance & Trade*, 51(5), 963–974. doi:10.1080/1540496X.2015.1061387
- Lara, J. A., De Sojo, A. A., Aljawarneh, S., Schumaker, R. P., & Al-Shargabi, B. (2020). Developing big data projects in open university engineering courses: Lessons learned. *IEEE Access : Practical Innovations, Open Solutions*, 8, 22988–23001. doi:10.1109/ACCESS.2020.2968969
- Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: Application of big data. *Journal of Big Data*, 4(1), 20. Advance online publication. doi:10.1186/s40537-017-0077-4
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 4. PMID:23074865
- Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10(7), 1497–1516. doi:10.1016/j.adhoc.2012.02.016
- Moore, J. (2017). Data Visualization in support of executive decision making. *Journal of Information, Knowledge, and Management*, 12, 125–138. doi:10.28945/3687
- Mousannif, H., Sabah, H., Douiji, Y., & Sayad, Y. O. (2014). From big data to big projects: A step-by-step roadmap. *Proceedings - 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014*, 373–378. doi:10.1109/FiCloud.2014.66
- Mousannif, H., Sabah, H., Douiji, Y., & Sayad, Y. O. (2016). Big data projects: Just jump right in! *International Journal of Pervasive Computing and Communications*, 12(2), 260–288. doi:10.1108/IJPC-04-2016-0023
- Olsson, N. O. E., & Bull-Berg, H. (2015). *International Journal of Managing Projects in Business*. *International Journal of Managing Projects in Business*, 8. 10.1108/IJMPB-09-2014-0063
- Poleto, T., De Carvalho, V. D. H., & Costa, A. P. C. S. (2017). The full knowledge of big data in the integration of interorganizational information: An approach focused on decision making. *International Journal of Decision Support System Technology*, 9(1), 16–31. doi:10.4018/IJDSST.2017010102
- Pondel, J., & Pondel, M. (2016). The concept of project management platform using BI and big data technology. *ICEIS 2016 - Proceedings of the 18th International Conference on Enterprise Information Systems*, 1, 166–173. doi:10.5220/0005834601660173
- Saheb, T., & Saheb, T. (2020). Understanding the development trends of big data technologies: An analysis of patents and the cited scholarly works. *Journal of Big Data*, 7(1), 12. Advance online publication. doi:10.1186/s40537-020-00287-9
- Saltz, J., Shamshurin, I., & Connors, C. (2017). Business Information Systems Workshops. *Lecture Notes in Business Information Processing*, 183–195. doi:10.1007/978-3-319-52464-1\_17
- Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. *Proceedings - 2016 IEEE International Conference on Big Data*, 2872–2879. doi:10.1109/BigData.2016.7840936

Salviati, M. E. (2017). *Manual do Aplicativo Iramuteq* (versão 0.7 Alpha 2 e R Versão 3.2.3). Retrieved from <http://iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>

Singh, N. (2019). Big data technology: Developments in current research and emerging landscape. *Enterprise Information Systems*, 13(6), 1–31. doi:10.1080/17517575.2019.1612098

Teixeira, A. A. C., & Queirós, A. S. S. (2016). Economic growth, human capital and structural change: A dynamic panel data analysis. *Research Policy*, 45(8), 1636–1648. doi:10.1016/j.respol.2016.04.006

Wang, W. (2020). Data analysis of intellectual property policy system based on Internet of Things. *Enterprise Information Systems*, 14(9-10), 1–19. doi:10.1080/17517575.2020.1712744

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: Innovation opportunities and challenges. *International Journal of Digital Earth*, 10(1), 13–53. doi:10.1080/17538947.2016.1239771

Yang, Z., Zhang, Z., Zhang, S., Wang, J., Lin, H., & Zeng, B. (2017). Article. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 2116–2122). IEEE. doi:10.1109/FSKD.2017.8393098