Journal of Management & Organization (2023), **29**, 697–723
doi:10.1017/jmo.2022.17

RESEARCH ARTICLE

Facilitating big-data management in modern business and organizations using cloud computing: a comprehensive study

Wenhao Qi[1], Meng Sun[2] and Seyed Reza Aghaseyed Hosseini[3,4]

[1]School of Economics and Management, Jilin Agricultural University, Changchun, Jilin 130118, China, [2]Center for Northeast Asian Studies, Jilin University, Changchun, Jilin 130012, China, [3]S P Jain School of Global Management, Sydney, Australia and [4]California Miramar University, California, USA
Author for correspondence: Meng Sun, E-mail: sunmeng58523@163.com

(Received 2 October 2021; revised 12 February 2022; accepted 4 March 2022)

Abstract
Modern digital life has produced big data in modern businesses and organizations. To derive information for decision-making from these enormous data sets, a lot of work is required at several levels. The storage, transmission, processing, mining, and serving of big data create problems for digital domains. Despite several efforts to implement big data in businesses, basic issues with big data remain (particularly big-data management (BDM)). Cloud computing, for example, provides companies with well-suited, cost-effective, and consistent on-demand services for big data and analytics. This paper introduces the modern systems for organizational BDM. This article analyzes the latest research to manage organization-generated data using cloud computing. The findings revealed several benefits in integrating big data and cloud computing, the most notable of which is increased company efficiency and improved international trade. This study also highlighted some hazards in the sophisticated computing environment. Cloud computing has the potential to improve corporate management and accountants' jobs significantly. This article's major contribution is to discuss the demands, advantages, and problems of using big data and cloud computing in contemporary businesses and institutions.

Key words: Big data; big-data management; cloud computing; modern management; organizations; security

Introduction

The globe is being inundated with data at a rate of 7 ZB per year, primarily from 'Internet of Things (IoT)' devices (Jamali, Bahrami, Heidari, Allahverdizadeh, & Norouzi, 2020). These data are dispersed across numerous devices, making it impossible to extract any meaningful relationships from them; conventional storage and processors are incapable of keeping up with this incredible velocity. Companies that are best equipped to make real-time business choices utilizing big-data solutions are expected to prosper. In contrast, those unable to adapt and exploit this change will progressively find themselves at a competitive disadvantage in the market and may collapse (Waga, 2013). In government, industry, and research, the demand to evaluate enormous quantities of data is growing. Data analysis is now considered the fourth paradigm in research (Hey, Tansley, & Tolle, 2009; Wang & Li, 2021). Users require tools to quickly and simply examine these data (Wang et al., 2017). Cloud computing has been widely adopted in the information technology (IT) sector, thanks to fully advanced cloud-computing business models, middleware technologies, and well-cultivated ecosystems (Wang, Ma, Yan, Chang, and Zomaya, 2018). Cloud computing is a platform that allows end-users worldwide to access a shared pool of resources on-demand over the internet. Physical servers in huge geo-distributed data centers host such shared pools of resources

(Chaudhary, Aujla, Kumar, & Rodrigues, 2018). Also, an impressive way to mitigate the overhead of computation is to offload the computing tasks to powerful devices, cloud, edge, or fog (Heidari, Jabraeil Jamali, Jafari Navimipour, & Akbarpour, 2020; Song, Cui, Li, Qiu, & Buyya, 2014).

Research motivation

The use of cloud technology for storing, processing, and analyzing big data is increasing. However, it also has some problems and challenges. Several studies have been done in this area. To motivate this study, some studies on the subject and their results are reviewed to find out the weaknesses of previous articles. Our goal is to fill in the gaps of previous articles. Table 1 provides information about these studies.

As reviewed, several reviews are conducted about this area. However, as Table 2 shows, no systematic reviews are provided. We intend to manage big modern data in an application group in a systematic review.

Cloud computing is an IT infrastructure that divides computing resources into service tiers and delivers them on demand. The service mode level is where the innovation is most visible, and the commercial value is achieved through fundamental operating features, including application hosting, resource leasing, and service outsourcing (Heidari & Navimipour, 2021; Sun, 2021). Increased communication, efficiency, and resource management are advantages of cloud computing, while big-data management (BDM) can fundamentally simplify internal and external connections. Adaptability, competitiveness, cost savings, and increased efficiency and profitability are the critical financial consequences of cloud computing. The most significant element in growing competitiveness in the creative industries is improved innovative capabilities. This research will look into the impact of cloud computing on BDM in businesses.

It serves as a benchmark for how IT management may affect the trajectory of the interaction between cloud computing and BDM deployment in the digital creative industries' innovative capabilities. It also adds to the body of information in the literature on performance enhancement in the digital creative sectors, which may be cited by data management and cloud-computing deployment innovation and management capabilities.

Therefore, in this systematic review, the following questions will be answered:

(1) In what areas can cloud technology for modern management be used? This question will be answered in Sections 'Research methodology and data statistics' and 'BDM in cloud.'
(2) What are the problems with using the cloud to manage big data? The answer to this question is in Sections 'BDM in cloud' and 'Results and discussion.'
(3) What can researchers do to improve the use of cloud technology for BDM? This question will also be answered in Sections 'Results and discussion' and 'Challenges.'

The remaining of the article is organized as follows. The second section 'Background' presents study background. The third section 'Research methodology and data statistics' describes the research method. The fourth section 'BDM in cloud' presents an overview of articles related to selective grouping. The fifth section 'Results and discussion' outlines the results and discussion, and the sixth section 'Challenges' restates some of the problems and challenges in BDM. The seventh section 'Future directions' is a guide to the future work of researchers. Eventually, the conclusions of the study are presented in the last section.

Background

Electronic gadgets and any computer-based (distributed) service are becoming progressively embedded in people's daily lives. Hence, in order to grow their incomings or enhance their services, businesses must analyze massive volumes of data (Amato & Moscato, 2016). Parallel

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Table 1. Some available review information

Paper	Goal and findings
Sahal, Khafagy, and Omara (2016)	*Goal*: Overviewing the progress in big-data and cloud-computing SLAs (i.e., representative applications such as BDAAs). *Findings*: (1) According to SLA in the cloud environment, the hybridization-based strategy, which combines several techniques, is superior to other ways for addressing SLA violations. (2) Some studies have presented effective scheduling strategies for BDAAs hosted on cloud resources that meet SLA constraints such as budget and timeline. (3) The notion of cross-layer SLA management can provide considerable benefits in preventing SLA violations for cloud-hosted BDAAs.
Ullah, Awan, and Sikander Hayat Khiyal (2018)	*Goal*: Categorizing enormous big-data resource management systems in terms of the cloud-computing environment. *Findings*: (1) For medium, small, and big data sets, Apache Flink surpasses competing resource management schemes in terms of processing performance. (2) Big-data applications often deal with large amounts of data. Although Apache Spark offers the required methods for fault tolerance, it has been observed to crash when dealing with big data sets. (3) To leverage scalable and distributed big-data applications, Spark MLlib and Flink-ML provide a number of ML techniques and tools. (4) Flink utilizes the Gelly library for graph processing algorithms such as PageRank, which supports native closed-loop iteration operators, making it a proper platform for large-scale graph analytics.
Zhu, Tan, Luo, and He (2018)	*Goal*: Explaining the history of cloud-enabled geological information services and the expectations from BDM and addressing the application demands and problems that BDM technologies encounter. *Findings*: With the continued advancement of big-data technologies in tackling issues associated with big geological data, such as the difficulty of explaining and modeling big geological data with some complex attributes, cloud-enabled geological information services will evolve in the coming years to become more intelligent and mature.
Almeida, de Aguiar Monteiro, de Lima, Hazin, and Escobar (2019)	*Goal*: Data integration, management, processing, and cloud-computing environment are all elements of big data that need to be addressed in this article. *Findings*: (1) Security and privacy are two of the most pressing concerns in the cloud-computing environment. (2) Several investigations have linked the adoption of standards to the deployment of progressively flexible solutions.
Inamdar, Raut, Narwane, Gardas, Narkhede, and Sagnak (2020)	*Goal*: Exploring and providing insights into the big-data analytics studies in different sectors. *Findings*: (1) The benefits of adopting big-data analytics, coupled with a lack of adequate research in the field, have motivated this study. (2) Founding that most of the studies were carried out in manufacturing and service.
Grander, da Silva, and Gonzalez (2021)	*Goal*: Presenting an overview of the techniques and technologies discussed in the literature to solve problems in their respective areas as a form of theoretical contribution. *Findings*: (1) Techniques of big-data analytics, ML algorithms, and technologies predominantly related to computer science and cloud computing are used on decision support systems. (2) The main areas where these techniques and technologies have been applied are logistics, traffic, health, business, and market.

database servers and cloud server technologies are two ways of managing a large quantity of data. Parallel database servers have been a huge success in both academics and industry since the early 1990s. Thanks to them, several apps that deal with huge amounts of data have fulfilled their performance and resource accessibility goals. Nevertheless, using a parallel database server is costly for a business. Furthermore, it necessitates acquiring a costly server and the availability of high-level talents within the firm to manage databases and servers (Hameurlain & Morvan, 2015).

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Table 2. Features of reviewed articles

Paper	Type	Classification	Future works	Year	Method description	Publisher
Sahal, Khafagy, and Omara (2016)	Survey	Yes	Yes	2016	No	Researchgate
Zhu et al. (2018)	Review	No	No	2018	No	Hindawi
Ullah, Awan, and Sikander Hayat Khiyal (2018)	Review	Yes	Yes	2018	No	Hindawi
Almeida et al. (2019)	Survey	No	No	2019	No	Researchgate
Inamdar et al. (2020)	Systematic Review	Yes	Yes	2020	Yes	Emerald
(Grander, da Silva, and Gonzalez, 2021)	Systematic Review	Yes	Yes	2021	Yes	Emerald

Because of the services it provides, cloud computing might be utilized as a foundation technology for a variety of technologies. Cloud computing is a novel generation of services aimed at providing access to information, apps, and data from any location at any time. Besides, Stergiou, Psannis, and Gupta (2020) introduced and detailed a new cloud-based system structure that relies on a unique federated learning scenario known as the integrated federated model – InFeMo. All cloud models with a federated learning scenario and additional technologies that might have been used in tandem were included in their model.

Big data and cloud computing have a close relationship. Big data in the cloud is a next-generation data-intensive platform that aims to provide rapid analytics across a flexible and scalable architecture. Cloud computing is a large computing capacity and infrastructure that allows storing and processing large data volumes, often known as big data. Besides, the emergence of big data has accelerated the growth of cloud computing. The cloud's distributed storage feature aids in the management of big data, while the parallel processing feature aids in collecting and analyzing large data (Agarwal & Srivastava, 2019).

Disk failures frequently cause outages in cloud-based services. Most of these failures are caused by electro-mechanical issues, which are nearly always visible in data utilized to monitor hard drives. The present procedures are reactive, which has an impact on the customer experience. Published work in disk failure prediction models is either outdated or barely 50–60% accurate (Pinheiro, Weber, & Barroso, 2007). Because the hard disk drives deployed in cloud systems are already tens of millions, proactively detecting problems and taking corrective action can provide considerable advantages (Ganguly, Consul, Khan, Bussone, Richards, & Miguel, 2016). In cloud computing, the existence of duplicated data is a critical issue. Data duplication is defined as the storage of the same data multiple times. Storage space is wasted when duplicated data are stored. Even though the cloud has a huge memory, duplicated information causes the large memory to be wasted, making data processing more difficult. As a result, deduplication has become more important in cloud data processing. Deduplication seeks to reduce storage costs. The cloud will become more profitable as a result of these savings. Deduplication is a key challenge when it comes to governing encoded info (Aslam & Swaraj, 2019).

Research methodology and data statistics

Using a systematic literature review technique to discover, choose, and analyze the particular field of study has recently received much attention (Esmailiyan, Amerizadeh, Vahdat, Ghodsi, Doewes,

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

& Sundram, 2021; Vahdat & Shahidi, 2020). The systematic literature review approach is being used to perform a survey because:

(1) comprehensively available in selected fields,
(2) reviewing relevant research, and
(3) prevent accidental or intentional omission of important research work to achieve the desired result.

It leads to the elimination of a set of studies that sufficiently represent the field of research. Nowadays, investigators utilize a variety of systematic literature review techniques (Petersen, Vakkalanka, & Kuzniarz, 2015; Vahdat, 2021). Figure 1 illustrates the systematic literature review method used in this paper. We first extracted the desired articles using keywords through this method. Then, by reviewing the titles and abstracts of the articles, we removed unused articles such as review articles, irrelevant articles, and duplicate articles. Finally, 23 articles were analyzed. The list of these articles is given in Table 3.

In this article, we used keywords such as 'big data management,' 'big data management AND cloud,' 'big data management AND cloud AND organization,' 'big data management AND cloud AND education,' 'big data management AND cloud AND healthcare,' 'big data management AND cloud AND business,' 'big data management AND cloud AND smart city,' etc.

We systematically reviewed our literature in databases such as the Scopus, Springer Online Journal Collection, Google Scholar, ACM Digital Library, IEEEXplore, WoS, and ScienceDirect. Figure 2 illustrates the articles obtained from these databases in the last 15 years according to the publication year.

According to Figure 2, researchers' attention has increased to the use of cloud technology in managing data from different applications such as organizations and offices in recent years. It is good to know what is the contribution of famous publications in these published articles. Figure 3 shows the contribution of each popular publication.

According to Figure 3, it is concluded that the articles we are interested in have not necessarily been published in well-known publications. However, other publications have a larger contribution to this research. Therefore, regardless of the publication, we will examine any research if relevant to our study's subject. By reviewing 220 articles found and reviewing their titles and abstracts, the desired grouping was selected as shown in Table 3.

BDM in cloud

Big data is created in different organizations and departments and should be stored, processed, and analyzed. Certainly, the use of cloud technology will be helpful to improve the management of these data. Studies on some applications are considered in this section. Therefore, articles will be studied in the grouping as shown in Figure 4.

Smart city

Big-data and cloud-computing analytics are critical components of smart city construction (Zhuang, Zhu, Huang, & Pan, 2021). They may help communities become more dependable, safe, healthy, and informed while also creating massive data for the public and commercial sectors. Because smart cities create massive volumes of streaming data from sensors and other devices, preserving and analyzing this massive real-time data generally necessitates a substantial amount of computer power. The majority of smart city solutions combine basic technologies such as computers, databases, storage, and data warehouses with modern technologies such as big-data analytics, artificial intelligence, real-time streaming data, machine learning (ML), and the IoT (Maroli, Narwane, & Gardas, 2021; Suresh et al., 2021).

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

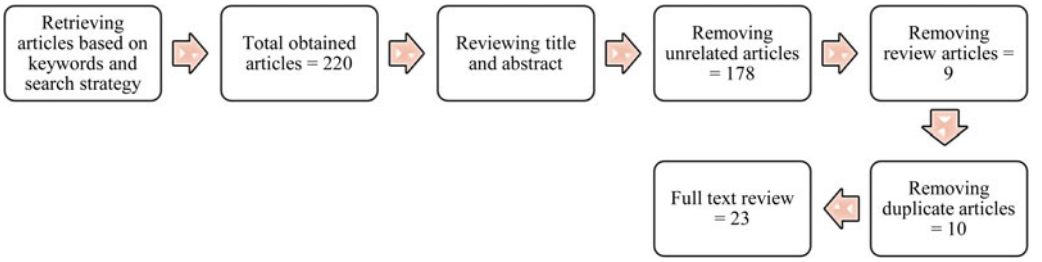

Figure 1. Research hierarchy.

In this section, 23 articles will be analyzed, six of which will be related to the smart city.

Sinaeepourfard, Krogstie, and Petersen (2018) created a hierarchical distributed data management infrastructure for a zero-emission community center in Norway. In the beginning, they described (from creation to consumption) the hierarchical distributed architecture capable of organizing the whole data life cycle levels. Afterward, they demonstrated that each cross-tier of the infrastructure (from IoT devices to cloud technologies) could handle various types of acquired data (containing recent, real time, and historical data). They described that fog-to-cloud data management (from distributed to centralized) has a great possibility to handle all data life stages (from creation to conception) concerning the data life cycle concepts. Also, they contributed to different smart city scenarios to demonstrate their proposed big-data architecture for smart cities. Also, in Gupta and Godavarti (2020), IoT data management utilized cloud and big-data technology to build a system that can manage the vast and rapidly expanding amount of data created by IoT devices. Its goal was to provide a more secure, scalable, fault-tolerant, and cost-effective environment for analyzing large data using cloud-computing services. A paradigm is presented in the suggested technique to effectively manage data supplied by IoT devices through Rest Application Programming Interfaces (APIs). The outcomes are given to show how the Rest API works throughout all nodes in a cluster using Javascript Object Notation (JSON) requests. The model was fed a request with a matching JSON payload. The transactions were added to the registered nodes with no need to add the payload again. A fresh batch was produced with all of the devices' readings. While retrieving the findings, the contents of the complete batch and all systems were retrieved, indicating the efficacy of the planned work.

Baek, Vu, Liu, Huang, and Xiang (2014) unveiled the Smart-Frame, a generic framework for managing large data sets in smart grids using cloud-computing technology. Their fundamental concept was to create three hierarchical layers of cloud-computing centers to handle information: top, regional, and end-user. The top cloud level provided a worldwide perspective of the architecture, while each regional cloud center was responsible for processing and maintaining regional data. Besides, they proposed identity-based cryptography and identity-based proxy re-encryption-based solution. Hence, not only does their suggested framework have scalability and flexibility, but it also has security characteristics. They created a proof-of-concept for our framework using a basic identity-based data confidentiality management system. Additionally, Kaseb, Mohan, and Lu (2015) demonstrated a system that employed the suggested resource manager to analyze large data from worldwide network cameras for video and image analysis. Investigations confirmed that using a resource manager can result in a cost reduction of 13%. Four analytic programs were employed throughout the studies, each representing a distinct workload regarding CPU and memory. In addition, the tests revealed that certain cloud instances were more cost-effective for various analytic procedures. Using multiple analytic programs at varying frame speeds, one study evaluated data streams from 1026 cameras concurrently for 6 hr. The study looked at 5.5 million pictures, totaling 260 GB of data. Besides, Park, Kim, Jeong, and Lee (2014) developed and tested the two-phase group categorization in a range of mobile device

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Table 3. Details of selected studies

Article	Publisher	Journal	*h*-index	Year	Cite	Authors
Baek et al. (2014)	IEEE	IEEE Transactions on Cloud Computing	1	2014	251	Joonsang Baek, Quang Hieu Vu, Joseph K. Liu, Xinyi Huang, and Yang Xiang
Park et al. (2014)	Wiley	International Journal of Communication Systems	1	2014	28	JiSu Park, Hyongsoon Kim, Young-Sik Jeong, and Eunyoung Lee
Kaseb, Mohan, and Lu (2015)	IEEE	International Conference on Cloud Computing and Big Data	1	2015	27	Ahmed S. Kaseb, Anup Mohan, and Yung-Hsiang Lu
Sreekanth, Rao, and Nanduri (2015)	Researchgate	International Journal of Advanced Engineering and Global Technology	2	2015	2	Sreekanth R., Golajapu Venu Madhava Rao, and Srinivas Nanduri
Huang et al. (2015)	POS	Proceedings of Science	1	2015	11	Jiayi Huang, Peipei Guo, Qiubo Xie, and Xiangbao Meng
Wang and Zhao (2021)	IEEE	IEEE 3rd International Conference on Cyber Security and Cloud Computing	1	2016	11	Ziqi Wang and Haihui Zhao
Celesti et al. (2016)	IEEE	International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)	1	2016	13	Antonio Celesti, Maria Fazio, Agata Romano, and Massimo Villari
Zuo (2017)	Atlantis Press	International Conference on Education, Management, Arts, Economics and Social Science	0	2017	–	Zuo Xiangli
Shan et al. (2017)	Springer	Proceedings of the 23rd International Conference on Industrial Engineering and Engineering Management	1	2017	4	You-cheng Shan, Chao LV, Qiu-ye Zhang, and Xin-yu Tian
Das et al. (2017)	Springer	Cluster Computing	1	2017	31	Amit Kumar Das, Tamal Adhikary, Md. Abdur Razzaque, Majed Alrubaian, Mohammad Mehedi Hassan, Md. Zia Uddin, and Biao Song
Thanigaivasan et al. (2018)	Researchgate	Recent Patents on Computer Science	1	2018	7	Vivekanandan Thanigaivasan, Swathi Jamjala Narayanan, Sriman Narayana Iyengar, and N. Ch
Yang (2018)	Web of Proceedings	Institute of Management Science and Industrial Engineering. Proceedings of 2018 International Workshop on Advances in Social Sciences	1	2018	1	Yi Yang

(*Continued*)

Table 3. (*Continued.*)

Article	Publisher	Journal	*h*-index	Year	Cite	Authors
Sinaeepourfard, Krogstie, and Petersen (2018)	NTNU	–	1	2018	16	Amir Sinaeepourfard, John Krogstie, and Sobah Abbas Petersens
Zhang et al. (2018)	Researchgate	International Conference on Management, Economics, Education, Arts and Humanities	0	2018	–	Bin Zhang, Bin Fang, Jianwei Yin, and Xiangyang Yu
Li (2019)	AEE Science	International Conference on Advanced Education, Service and Management	1	2019	2	Yangyang Li
Terrazas, Ferry, and Ratchev (2019)	Elsevier	Computers in Industry	1	2019	2	German Terrazas, Nicolas Ferry, and Svetan Ratchev
Chen and Dou (2020)	Springer	International Conference on Big Data Analytics for Cyber-Physical-Systems	0	2020	–	Jianhua Chen and Huili Dou
Gupta and Godavarti (2020)	IGI	International Journal of Software Innovation (IJSI)	1	2020	4	Sangeeta Gupta and Raghuram Godavarti
Munir et al. (2020)	ProQuest	Journal of Quantum Computing	0	2020	–	Rizwan Munir, Yifei Wei, Rahim Ullah, Iftikhar Hussain, Kaleem Arshid, and Umair Tariq
Yan and Nanyun (2020)	Conferences	E3S Web of Conferences	0	2020	–	Xu Yan and Xiao Nanyun
Jain (2020)	SSRN	International Journal of Computer Engineering and Technology	0	2020	–	Nupur Jain
Xiaona (2021)	IOP	Journal of Physics: Conference Series	1	2021	1	Ma Xiaona
Ionescu and Andronie (2021)	ProQuest	SHS Web of Conferences	1	2021	1	Luminița Ionescu and Mihai Andronie

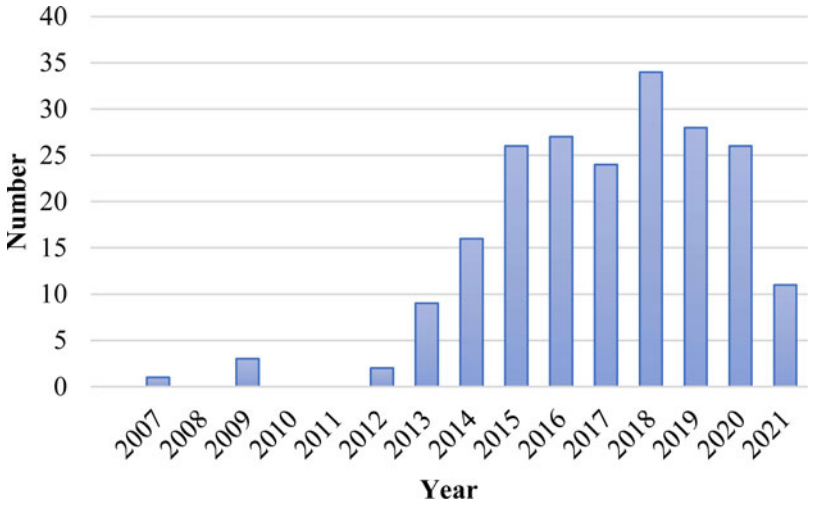

Figure 2. Distribution of studies by year.

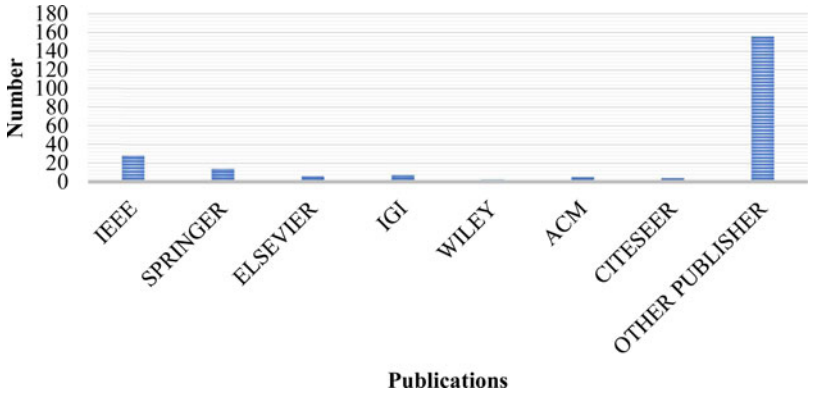

Figure 3. Percentage of articles in each publication.

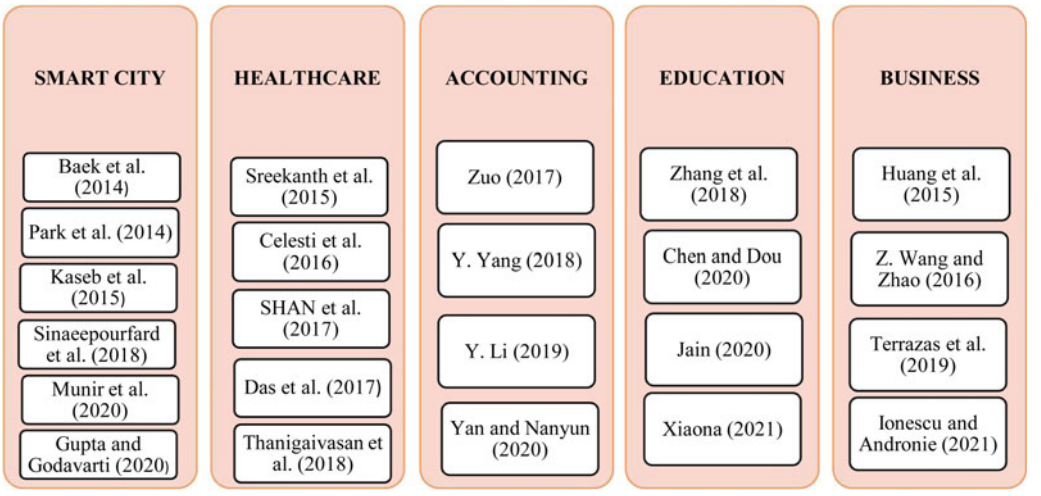

Figure 4. Grouping selected journals.

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

distributions. Previous investigations that used arbitrary cut-off thresholds were ineffective in mobile cloud systems, which had a high level of instability. The recommended approach created a two-phase grouping by merging groups from entropy-based grouping with displaying group similarity. Even when the distribution of mobile devices varies, the algorithm correctly produces two-phase groups, according to the testing outcomes. When it came to sustaining reliable massive data processing and managing dependable resources, their algorithm beat standard grouping approaches.

Munir, Wei, Ullah, Hussain, Arshid, and Tariq (2020) described a cloud-computing-based smart grid system that incorporates a big-data strategy. Data source, storage/processing, transmission, and analysis were the four levels that make up the architecture. A case study was created using a data set from three cities in the Pakistani region and two cloud-based data centers. High load (on data centers) and network latency, according to the research, may impair overall efficiency by generating a reaction time delay. They argued that having a local data center might help minimize data load and network delay. For both customers and service suppliers, the provided paradigm may be useful in achieving sustainability, reliability, and cost-effectiveness in the power grid.

To conclude and summarize the articles related to the smart city, Table 4 provides some details and features of these studies.

Healthcare

The world's population is growing, expecting more effective treatments and a higher overall quality of life. It is putting more strain on healthcare (Simpson, Farr-Wharton, & Reddy, 2020). As a result, healthcare continues to be one of the world's most pressing social and economic issues, requiring newer and more developed solutions from technology and science (Aceto, Persico, & Pescapé, 2020; Chiuchisan, Costin, & Geman, 2014; Omanović-Mikličanin, Maksimović, & Vujović, 2015). The following research provides a solution to these challenges. Five articles related to this topic are as follows.

In Thanigaivasan, Narayanan, Iyengar, and Ch (2018), the heart disease data set was used for analysis. The data set was used in several tests to assess the performance of classification algorithms, and support vector machine (SVM) was shown to outperform the others. In the case of huge data, SVM was discovered to have a long processing time. Thus, the large-scale data set was classified using parallel SVM-based categorization. The parallel SVM substantially decreased the processing time, properly classifying the data. Besides, Celesti, Fazio, Romano, and Villari (2016) spoke about an open archival information system -based hospital information system that can manage large amounts of data in a cloud-computing environment. They explored two alternative executions of archival storage sub-components based on MySQL and MongoDB, respectively, regarding the health level 7 v3 standard. Studies demonstrated that MongoDB was an excellent candidate for implementing an archival storage sub-component capable of handling large amounts of data. In reality, while SQL is the most widely used technology for archival storage in hospital information systems worldwide, it cannot meet the new difficulties posed by cloud-based hospital information systems and big health data. In comparison with MySQL, MongoDB makes it easier to retain health level 7 documents with minimum processing work.

Sreekanth, Rao, and Nanduri (2015) looked at how MongoDB may handle and analyze large data in electronic health records systems on the cloud. Afterward, they explored creating an electronic health records system using MongoDB, an NoSQL database. Because electronic health records are projected to grow in popularity, a system based on NoSQL is essential. Document-based JSON files can be used to create electronic healthcare-records systems. Systems based on NoSQL outperform SQL-based systems. Additionally, Shan, Chao, Zhang, and Tian (2017) discussed the meanings of big data and cloud computing and the state of health management studies in the country and overseas. It also explained the data methodology and

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Table 4. Details of the analyzed articles of the smart city group

Article	Motivation	Finding	Advantage
Park et al. (2014)	• Because of the increased computational capabilities of mobile devices, distributed processing methods such as MapReduce are being applied to them. • On the other hand, mobile devices have some issues, including movement and usage. • As a result, they developed a grouping approach based on use and mobility rates to address these serious issues with mobile devices.	• When it comes to sustaining consistent massive data processing and managing dependable resources, its method beats standard grouping strategies.	• Reducing the overhead of group management.
Kaseb, Mohan, and Lu (2015)	• Demonstrating a cloud resource manager to lower the total cost of processing several video and image feeds from network cameras.	• The proposed resource manager can lead to a 13% reduction in the overall analysis cost.	• More cost-effective
Sinaeepourfard, Krogstie, and Petersen (2018)	• Developing BDM architecture proposal for the zero-emission neighborhoods research centers.	• Each cross-layer (from IoT devices to cloud technologies) in the structure may manage multiple sorts of acquired data (such as recent, real-time, and historical data).	• Decreasing communication latencies for critical or real-time services • Reducing the traffic of network data. • Using various policies (for instance, data aggregation, data filtering, data securities, etc.).
Gupta and Godavarti (2020)	• Better management of data generated by IoT devices through the Rest API.	• The proposed work is effective.	• High security • Fault-tolerant • High scalability • Cost-effective
Munir et al. (2020)	• Researchers focused on the smart grid's relevance, characteristics, possibilities, and difficulties and presented an infrastructure for wireless big-data analysis in research published by Cao, Lin, Wan, Song, Zhang, and Wang, (2016) and Okay and Ozdemir (2016). • They refined this generic design more and simulated it using a data set of three cities in the Pakistan region, inspired by it.	• Heavy load (on data centers) and network latency can reduce overall efficiency by causing response times to be delayed. • Having a local data center might help to minimize network latency and data load.	• High reliability • High sustainability • Cost-efficiency

essential technologies before going over the monitoring data transfer procedures. It also highlighted a novel pattern that employs a cloud-based warning data platform as a carrier to provide all types of early warning services to hospitals, communities, families, and other subscribers in the health management system. Furthermore, Das et al. (2017) created a global and local cloud confederation architecture, dubbed FnF, for performing heterogeneous large healthcare data processing demands from consumers. FnF uses fuzzy logic to make an appropriate selection decision for target cloud data center(s). In choosing a federated data center(s), the FnF trades off between user application Quality of Service (QoS) and cloud provider profit. Furthermore, FnF improves its decision accuracy by utilizing multiple linear regression to properly estimate the resource needs for massive data processing tasks. Numerical and empirical assessments were used to validate the suggested FnF model. In comparison with modern techniques, simulation outcomes demonstrated the efficacy and efficiency of the FnF model.

Everything obtained in this section is summarized in Table 5. Some features of the articles are listed in this table.

Accounting

In the big-data sector, cloud computing and large accounting data are combined to produce a cloud-accounting application framework that emphasizes spatial accessibility, security, distribution, and changing the accounting data condition. Confronted with a tidal wave of economic expansion, administrative agencies will begin to use cloud accounting, which will show considerable promise in these sectors (Li, 2021; Nosratabadi, Mosavi, Shamshirband, Kazimieras Zavadskas, Rakotonirainy, & Chau, 2019). The following four articles are related to this section and explain some challenges and benefits of this data management in the cloud.

The growth of agricultural firms is inextricably linked to the growth of the local agricultural sector. Nonetheless, the utilization of cloud accounting in agricultural businesses is restricted, and its application in comprehensive budget management is constrained, hindering agricultural businesses and the economy from docking efficiently (Yan & Nanyun, 2020). So, they thought that agricultural businesses might benefit from the benefits of big-data and cloud-accounting platforms by developing a more information-based comprehensive budget management system, which would help them strengthen their core competitiveness. Besides, Li (2019) discussed the importance of cloud computing and big data in management accounting and the possibilities and difficulties that management accounting education faces in the big-data era. Accordingly, they discussed how to incorporate management accounting and cloud-accounting systems efficiently, based on their extensive teaching experience, in order to support the fast growth of management accounting education. Also, Zuo (2017) discussed the impacts cloud accounting and big data have on an enterprise's overall budget management. The system then develops a framework for the company's complete budget management system, which optimizes budget enforcement, budget modification, budgeting, and budget evaluation operations, leading to rational resource allocation. Big data provides more extensive and accurate data assistance with new opportunities and directions for comprehensive budget management. He illustrated the impact of big data on comprehensive budget management and proposed a systematic framework for setting budgeting, strategic goals, enforcing budgets, and evaluating budgets to attain an appropriate allocation of company resources.

Yang (2018) put forward the methods to solve the dilemma of data standards from the three principles of standard data formulation, formulation ideas, and specific recommendations. From the seven aspects of technical means and management methods, he put forward the idea of solving security dilemmas. Therefore, enterprises should strengthen the application of cloud-accounting technology to meet enterprise development needs under the era of big data and promote better development of enterprises. The results of the analyzed articles in this section are summarized in a table. Table 6 shows these details better.

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Table 5. Details of the analyzed articles of the healthcare group

Article	Motivation	Finding	Advantage
Sreekanth, Rao, and Nanduri (2015)	• SQL is not built to deal with large amounts of data. Traditional database management systems have constraints in terms of scalability and infrastructure costs.	• NoSQL-based systems perform better than SQL-based systems.	• Scale support
Celesti et al. (2016)	• Hospital archival storage systems are usually built on outdated relational database management systems, making managing patient data, particularly in the health level 7 format.	• For hospitals, archival storage for big-data processing is easier than standard archival storage systems in a scaled situation.	• Able to manage big data. • When compared to MySQL, MongoDB simplifies storing health level 7 documents with fewer parsing tasks. • MongoDB, similar to other NoSQL databases, can flexibly save unorganized data and retrieve large amounts of data quickly.
Shan et al. (2017)	• Health management and IT of the competent departments to promote the development of cloud computing and big data.	• Using cloud computing and big data to improve the warning data ability of health management.	• Identifying and dealing with the mass of data generated by health management, optimizing the monitoring data transmission process, building cloud services data platform, and exploring the new path of warning data to provide effective solutions of health management for families, communities, and hospitals.
Das et al. (2017)	• Performing a variety of user-generated large healthcare data processing demands.	• In comparison with modern techniques, simulation findings show how effective and efficient the FnF model is.	• Increasing decision accuracy
Thanigaivasan et al. (2018)	• Meeting the challenge of rapid data growth in the healthcare domain.	• SVM was used to classify a CVD data set more efficiently than other classification techniques.	• Reducing the execution time • Classification of the data accurately

Table 6. Details of the analyzed articles of the accounting group

Article	Motivation	Finding	Advantage
Zuo (2017)	• Using cloud accounting to demonstrate scientific mass communication and government management.	• Big data provides more extensive and accurate data assistance for comprehensive budget management, supplying new opportunities and directions.	• More effective • Accurate and dynamic comprehensive budget management • Realizing their strategic goals
Yang (2018)	• Designing methods to address the issue of data standards.	• From the seven aspects of technical means and management methods, the paper puts forward the idea of solving security dilemmas.	• Realization of information management
Li (2019)	• Solve management accounting challenges.	• Using a large teaching background shows how to properly combine management accounting and cloud-accounting systems to encourage the fast growth of management accounting education.	• The opportunities and challenges facing management accounting training in the big data age are mentioned.
Yan and Nanyun (2020)	• Providing services for accounting and management of companies through cloud accounting based on network IT.	• Agricultural companies may utilize the benefits of big-data and cloud-accounting platforms to create a more information-based comprehensive budget management system, which will help them strengthen their core competitiveness.	• Accelerate the realization of accounting information • Reducing the cost of business operation • Increasing business efficiency

Education

Learners' learning has shifted from a single conventional instruction style to a composite learning model of classroom teaching and network learning (Anshari, Alas, & Guan, 2016). Due to the growth of network technology, their learning time has grown more diversified and dispersed. The conventional teaching approach has failed to fulfill learners' various learning demands. In light of this, online education based on an online training platform with features such as autonomy, customization, and interaction has emerged as a necessary component of modern training (Wang & Zhao, 2021). Four articles in this grouping will be analyzed as follows.

Jain (2020) offered cloud data security techniques and strategies to assure protection by reducing risks and hazards to a minimum. They addressed offering data security, network security, and privacy-preserving for cloud-computing security concerns. The suggested technique allows academic institutions to safely and efficiently retain data in the cloud. It proposed encryption and compression-based solution to the challenge of massive data security concerns. The outcomes of the experiments revealed that the suggested approaches outperform other systems in terms of efficiency and accuracy. Besides, Jianhua Chen and Dou (2020) concentrated on studying university education and teaching management's informatization approach in the big-data and cloud-computing era. The cloud computing and big-data age first looked at the current status of university education and teaching management informatization. Afterward, it analyzed and constructed a set of info management systems using an implemented parameter setting and collaborative filtering algorithm. Eventually, it described and addressed it from various perspectives. In the cloud computing and big data, the executing measures of university education and teaching management informatization sought to provide reference materials for related previous studies.

Zhang, Fang, Yin, and Yu (2018) created a university P.E. cloud platform management system based on a big-data analysis methodology and blockchain technology. It had a positive influence on the current state of university P.E. and the quality of teaching. The management system integrated traditional sports health data analysis, education management, and big-data analysis for the first time. It sought to apply new blockchain technology to increase data security, reliability, and reuse (Dehghani et al., 2020, 2021). Additionally, Xiaona (2021) looked at the relevance of informatization in education and teaching management in the big-data and cloud-computing era and offered strategies to build info in education and teaching management to aid the pertinent. To summarize, the advancement of educational and instructional management information can greatly improve pupil efficiency, effectiveness, and inner capability. To accomplish information management, it is important to include the notion of innovative instructional techniques and models and an innovative training system, environment, and philosophy into management. Universities and colleges should set a clear goal in line with their own advancement education teaching information, create a perfect information management system, continuously improve the education teaching management informatization level, establish scientific solutions, and instill high-quality talents for society in the big-data and cloud-computing era.

In this section, a table is created that describes the characteristics of the analyzed articles. Table 7 contains this information.

Business

In small and medium businesses, cloud-based architecture adds a whole novel dimension to data and insight sharing (Kars-Unluoglu & Kevill, 2021; Xiang, Zhang, & Worthington, 2018). Small and medium businesses may not have the resources or desire to run their own big-data architecture (Lan & Unhelkar, 2015). The cloud platform will help manage the data they generate. The following articles illustrate this point (Chen, Gao, & Ma, 2021a; Chen & Sivakumar, 2021).

Ionescu and Andronie (2021) aimed to explain and illustrate the difficulties regarding financial consequences resulting from BDM and cloud-computing's effect in the digital world. They

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Table 7. Details of the analyzed articles of the education group

Article	Motivation	Finding	Advantage
Zhang et al. (2018)	• Development of Traditional university physical education (P.E.) in the context of internet and big data.	• Through the actual use and testing, the application process of the system in teaching is gradually improved and matured.	• Improving the quality of teaching effects • Improving the current state of the university • Better security, reliability
Jianhua Chen and Dou (2020)	• Analyzing the info strategy of university education and teaching management in the big-data and cloud-computing age.	• Colleges and universities should thoroughly understand the impact of the big-data and cloud-computing age on education and teaching management.	• Improving the informatization level of education teaching management.
Jain (2020)	• Creating data security in small and medium educational institutions.	• The results of the experiments demonstrate that the suggested approaches perform better than previous systems in terms of accuracy and efficiency.	• High security
Xiaona (2021)	• Innovating and optimizing college education and teaching management, changing traditional models and methods, building education and teaching management system, and improving management level.	• Analyzing the relevance of info of education and teaching management, and proposing approaches to create info of education and teaching management to assist the relevant stakeholders in the big-data and cloud-computing age.	• Improving the efficiency and quality of students and internal potential.

employed a combination of qualitative and quantitative investigation to identify the benefits of using BDM with a direct favorable influence on corporate performance. Their research looked into the financial implications of cloud-computing and digital solutions for businesses in the digital era and the impact of cloud technology usage on business growth. There are several benefits to integrating cloud computing and big data, but the most significant is increasing company efficiency and improving the global economy. Additionally, Terrazas, Ferry, and Ratchev (2019) demonstrated a new big-data strategy and analytics architecture for the cloud administration and analysis of machine-produced data. It combined open source technology with the use of elastic computing to create a system that can be modified to and deployed on a variety of cloud-computing platforms. The outcome is a distributable, versatile, and scalable solution that allows for easy incorporation of technologies that can adapt to various manufacturing settings and cloud-computing suppliers. It allowed for lower easier deployment, infrastructure costs, and on-demand accessibility to a nearly limitless pool of storage, computing, and network resources.

Huang, Guo, Xie, and Meng (2015) merged e-commerce with conventional business models utilizing network technology, database technology, cloud-computing technology, and marketing management technology to create an incorporated cloud services platform for advanced livestock

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

marketing management to meet the actual necessities of contemporary livestock marketing management. The platform combines e-commerce and conventional business models to supply outsourcing services for livestock enterprises, such as customer relationship management, e-commerce, inventory management, and more. It assists livestock enterprises in selling products and enhancing production management levels by incorporating e-commerce and conventional business models. Promoting traditional to contemporary transformations, improving management levels, increasing competitiveness, and promoting economic advantages benefit the livestock sector. Furthermore, Wang and Zhao (2016) provided experimental research on leveraging big data in cloud computing to optimize business processes. The study focused on a large-scale Chinese private company that aspires to be a worldwide player in the manufacturing business. The completed investigation was based on real data obtained from the collaborating partner. The fundamental outcomes of their study were as follows: the attempts to use big data differed according to the operating levels; adopting cloud-computing solutions for the Chinese private sector was exploratory due to some constraints. The outcomes revealed the current cloud computing and big-data deployments in Chinese private enterprises.

Four articles are reviewed in this section; the most important points of this study are summarized in Table 8.

Results and discussion

In the previous section, 23 articles were studied in smart city, education, health, and business. The results of the studies were expressed in tables. According to articles in the previous section, managing a large amount of big data, which may involve data selection, monitoring, deployment, and analysis, is unquestionably difficult (Wang, Wang, & Li, 2021). More crucially, real-time data processing is generally necessary with the smart grid. Any delay in the system might have significant consequences, which must be prevented as much as feasible (Baek et al., 2014).

Some frameworks, databases, and other research information were also extracted. Some have used regression or ML or fuzzy logic to manage big data (Liu, Zhang, & Lu, 2020; Zhong, Fang, Liu, Yuan, Zhang, & Lu, 2021). In addition to the proposed framework, the MapReduce and Hadoop frameworks are often used. MapReduce is a distributed computing system for dealing with huge unorganized data collections. To put it another way, MapReduce splits input files into pieces and processes them in stages. Hadoop, an open-source version of MapReduce, was also presented. Massive data sets may be processed with Hadoop clusters. Large data sets may be processed utilizing the MapReduce architecture and 'cloud' resources. Cloud computing provided a wide range of applications and decreased IT expenses, resulting in a substantial increase in efficiency. Remote control or software virtualization is a basic and general description of cloud computing (Roshandeh, Poormirzaee, & Ansari, 2014). The variety of big-data analytics platforms available in the cloud makes emergency decision-making difficult for solution architects, software developers, and infrastructure managers (Puthal, Nepal, Ranjan, & Chen, 2016).

The MongoDB database is also used. Actually, similar to other NoSQL databases, MongoDB can flexibly hold unorganized data and quickly retrieve large amounts of data (Celesti et al., 2016). Although SQL has numerous benefits, such as transaction security, NoSQL systems may be built for less than 10 times the cost of SQL systems (Sreekanth, Rao, & Nanduri, 2015). Table 9 illustrates the summary of the articles.

Challenges

Because cloud computing provides the platform, software, and infrastructure as a service (Ayala, Vega, & Vargas-Lombardo, 2013; Shahid, Ashraf, Ghani, Ghayyur, Shamshirband, & Salwana, 2020) and hosts apps through computer resources, platforms, or the internet, it faces a number of problems. Compromises in service quality, security, privacy, virtualization, scalability, integrity,

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Table 8. Details of the analyzed articles of the business group

Article	Motivation	Finding	Advantage
Huang et al. (2015)	• Promoting the distribution of knowledge, enhancing the capacity to employ IT, and extending the market for livestock goods to speed up the growth of the livestock sector.	• The platform utilizes the internet to offer outsourcing services for livestock businesses, such as e-commerce, inventory management, customer relationship management, and more. It assists livestock businesses in selling goods and enhancing production management by incorporating e-commerce and conventional business models.	• Improving the conventional transition to contemporary • Increasing the degree of management • Increasing competitiveness
Z. Wang and Zhao (2016)	• Improving the business process in computing using big cloud data.	• Cloud-computing and big-data deployments in the Chinese private sector.	• Enhancing business procedures • Utilizing two methods for study, containing A-Rule and performance matrix
Terrazas, Ferry, and Ratchev (2019)	• Analyzing the massive amounts of data created necessitates highly scalable systems in terms of both computer power and network capacity.	• An elastic computing model was established as part of the data analytics architecture. As a result, infrastructure costs are reduced, and deployment difficulties are minimized.	• Decreasing the expenses of infrastructure • Reducing the problem of deployment • Offering on-demand accessibility to an almost limitless amount of computer power • Resources for network and storage
Ionescu and Andronie (2021)	• Highlighting the difficulties in question regarding their financial consequences as a result of BDM and cloud-computing's effect in the digital world.	• Illustrating the financial consequences of cloud computing and digital solutions for organizations in the digital era are presented, and an impact assessment of cloud technology adoption in business development.	• Improving the firms' productivity • Enhancing the international economy

Table 9. Tools used in articles

Category	Article	Database used	Algorithm used	Architecture	Framework	Platform
Smart city	Baek et al. (2014)	–	–	–	Smart-frame	–
	Park et al. (2014)	–	–	–	MapReduce	–
	Kaseb, Mohan, and Lu (2015)	–	–	–	–	CAM2
	Sinaeepourfard, Krogstie, and Petersen (2018)	–	–	Based on fog-to-cloud	–	–
	Gupta and Godavarti (2020)	–	–	Suggested architecture	Hadoop, MapReduce	–
	Munir et al. (2020)	–	–	–	Proposed framework	–
Healthcare	Sreekanth, Rao, and Nanduri (2015)	MongoDB	–	–	Hadoop	–
	Celesti et al. (2016)	MongoDB	–	–	–	–
	Shan et al. (2017)	–	–	–	–	Proposed platform
	Das et al. (2017)	–	–	FnF	–	–
	Thanigaivasan et al. (2018)	–	Parallel SVM	–	–	–
Accounting	Zuo (2017)	–	–	–	Proposed framework	–
	Yang (2018)	–	–	–	–	Proposed platform
	Li (2019)	–	–	–	–	Proposed platform
	Yan and Nanyun (2020)	–	–	–	Proposed framework	–
Education	Jain (2020)	–	–	–	–	Cloud platform based on blockchain
	Jianhua Chen and Dou (2020)	–	Collaborative filtering algorithm	–	Cloud-based data analytics framework	–
	Jain (2020); B. Zhang et al., (2018)	–	–	–	–	Proposed platform
	Xiaona (2021)	–	–	–	–	–

(*Continued*)

Table 9. (*Continued.*)

Category	Article	Database used	Algorithm used	Architecture	Framework	Platform
Business	Huang et al. (2015); Ionescu and Andronie (2021)	–	–	–	–	Cloud services platform
	Z. Wang and Zhao (2016)	–	–	–	Proposed framework	–
	Terrazas, Ferry, and Ratchev (2019)	CouchDB	–	Suggested Architecture	cloud-based framework	–
	Ionescu and Andronie (2021)	Eurostat, Intuit and Software Advice databases	–	–	–	–

Table 10. Abbreviation table

Abbreviation	Definition
API	Application programming interface
BDM	Big-data management
IoT	Internet of Things
IT	Information technology
JSON	Javascript object notation
ML	Machine learning
OAIS	Open archival information system
QoS	Quality of service
SLA	Service level agreement
SVM	Support vector machine

and data debugging problems are among the concerns (Cheng, Shojafar, Alazab, Tafazolli, & Liu, 2021; Zhang, Chen, & Susilo, 2020). Never before have distributed storage and data management systems had to deal with problems such as data quantities and processing throughput related to the rise of big data to such a degree. Cloud storage systems are still in their infancy and are continuously evolving (Chen, Liu, Xiang, & Sood, 2021b). Until now, they have mostly concentrated on the requirements of commercial applications to deliver basic functionality dependably and securely (Shen, Zhang, Wang, Guo, & Susilo, 2021). Implementing data-intensive applications in the cloud at scale necessitates addressing the following issues.

- In many situations, streaming data transfers might be unreliable. On a daily basis, data sources create petabytes to terabytes of data (Hu, Wen, Chua, & Li, 2014). Real-time computing has become a big issue due to the collected volume (Puthal et al., 2016).
- Data staging is one of the most critical issues that must be addressed because data from sensors, mobile phones, and social networking sites are diverse. They lack any specific structure. In other words, sometimes the data accessible to analyze are unorganized data such as videos, text, and so on, necessitating extra work in cleaning and converting such data for processing, making the process sluggish and inefficient (Agarwal & Srivastava, 2019).
- Although cloud computing brings ease to businesses and individuals due to its structural qualities, it also unavoidably brings security threats from the computer network environment, creating a danger to the security of archived information resources (Shamshirband, Fathi, Chronopoulos, Montieri, Palumbo, & Pescapè, 2020; Sun, 2021).
- Manufacturing has become much easier because of the advent of sensor technologies that allow machines to communicate and gather data (Lee, Lapira, Bagheri, & Kao, 2013). As a result, the industrial information system has a huge problem figuring out utilizing and organizing massive data to help make better decisions (Li, Song, & Huang, 2016).
- One of the present data management problems is to deliver a service with no data loss and minimal throughput latency. Nevertheless, even after activating and incorporating a cloud management system, servicing all data streams and transactions remains a challenge (Hussien & Sulaiman, 2016).

Future directions

Although much research is done on the BDM of modern cloud systems, issues should be addressed. The following are important suggestions for the future:

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

- The urge to close the gap between data gathering and business action is growing. For instance, a shop could want to base next week's promotions on this week's information. It is desired for online shops to take action based on data even more rapidly. Available methods rely on log-based streaming, shipping, and other extracts, transform, and load approaches. However, this discipline is still in its early stages of growth (Chaudhuri, 2012).
- Despite the fact that enforcing service level agreements (SLAs) is a difficult undertaking, numerous academics have worked to build systems that might ensure that various services' QoS needs are met. In cloud computing, many ways to SLA violation have been presented. Even though resource allocation management is utilized to select appropriate resources for provider profit, cloud client demands, and cloud-hosted big-data analytic applications, it has not been properly examined (Sahal, Khafagy, & Omara, 2016).
- Scholarly data are a massive data repository that is constantly updated and contains a wide range of data. Hence, it is sometimes referred to as 'big scholarly data.' The analysis and display of these data may be used to create various applications (Hu et al., 2021). Difficulties and limits occur at every level of the data analytics procedure, particularly about big scholarly data platforms. Specific elements of this platform are undergoing study, which must be combined in order to build a comprehensive system (Khan, Shakil, & Alam, 2016).
- As the popularity of cloud-computing settings grows, so do the safety concerns that arise from this technology's adaption. As a result, there is necessary to invest in comprehending the loopholes, problems, and components that are vulnerable to attacks in cloud computing and developing a platform and architecture that is less vulnerable to assaults (Jain, 2020).
- Due to the abundance of wearable gadgets, smart sensors, smartphones, and other connected devices (Yi, 2021), fog/IoT will become the most researched subject in the subsequent decade (Heidari et al., 2020). As a result, data processing applications will most likely be deployed in a distributed manner. Nevertheless, sending all of the data to cloud data centers for processing is inefficient. It might result in unnecessary network, transmission, or bandwidth overhead across the system and increased data center energy usage. Hence, energy-efficient software solutions that can handle and analyze data at the fog/edge level must be created to minimize energy usage and improve the performance of time-critical applications (Yang et al., 2021). Additionally, multi-tiered resource management across the fog nodes, cloud data center, and mobile devices will aid in meeting the SLA need (Bagheri, Nurmanova, Abedinia, Naderi, Ghadimi, & Naderi, 2018; Islam & Buyya, 2019).
- Because the entire data cannot be transferred or processed, new techniques for filtering big data for processing must be created. Analyzing various data types attracts a wide range of studies (Anuradha & Bhuvaneshwari, 2014).
- The creation of a benchmark suite aimed at determining the highest throughput through configuration optimization would be a promising future study topic (Ullah, Awan, & Sikander Hayat Khiyal, 2018).

Conclusion

Companies are confronting issues such as optimizing resource allocations, cost control, managing quick storage growth necessities, coping with dynamic concurrency requests, and the lack of underlying infrastructures that can dynamically allocate the needed computing and storage resources for big data. As a result, the greatest answer is for a company to adopt new technology. Cloud computing is one such domain that has a major influence on how large data are handled, deployed, and consumed. Modern management is built in this work employing cloud and big-data technology to produce a system capable of handling the vast and rapidly expanding diversity of data-produced devices. Papers were thoroughly reviewed in this study. Hence, adopting the

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

data deduplication idea in the cloud has enabled users to reduce large data memory needs, lowering storage costs efficiently. Cloud computing is a prospective computing utility paradigm for delivering IT services to lower user costs. On the other hand, cloud computing is insecure. Attackers may penetrate the SaaS layer on cloud computing, exposing sensitive data and opening the door to a new form of hazardous assault. In addition, cloud-based big-data analytics has become a prominent study subject, posing new problems across the data processing life cycle, from data collection through integration and analytics to data security and privacy. These problems necessitate a novel system structure for data collecting, transmitting, storing, and large-scale data processing, replete with data privacy and security safeguards. Nevertheless, some progress has been made in this area. This paper attempts to create a more secure, scalable, fault-tolerant, and cost-effective environment for analyzing big data in companies using cloud-computing services. English sources are used in this study. There may be other valuable resources in other languages that are not listed here. In addition to the keywords we are looking for, there may also be other useful articles that are not selected through our selected keywords. In conducting this study, we have tried to use all sources without bias and justice. However, it is natural that some sources are inadvertently left out or, due to the diversity of research in this field, it is impossible to refer to them in this study. There may also be valuable resources other than English losses, which we ignore in this review.

Finally, the abbreviations used in the article are described in Table 10.

Acknowledgements. The authors acknowledge Scientific research project of Jilin Provincial Department of Education (JJKH20210377SK).

Conflict of interest. The authors declare no conflict of interest.

Data availability statement. All data are reported in the paper.

References

Aceto, G., Persico, V., & Pescapé, A. (2020). Industry 4.0 and health: Internet of Things, big data, and cloud computing for healthcare 4. *Journal of Industrial Information Integration*, *18*, 100129.

Agarwal, M., & Srivastava, G. M. S. (2019). 'Big' data management in cloud computing environment. In *Harmony search and nature inspired optimization algorithms* (Vol. *741*, pp. 707–716). Springer.

Almeida, W. H. C., de Aguiar Monteiro, L., de Lima, A. C., Hazin, R. R., & Escobar, F. (2019). Survey on Trends in Big Data: Data Management, Integration and Cloud Computing Environment.

Amato, F., & Moscato, F. (2016). *Automatic cloud services composition for big data management*. Paper presented at the 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA).

Anshari, M., Alas, Y., & Guan, L. S. (2016). Developing online learning resources: Big data, social networks, and cloud computing to support pervasive knowledge. *Education and Information Technologies*, *21*(6), 1663–1677.

Anuradha, D., & Bhuvaneshwari, S. (2014). A detailed review on agent-based computing in hybrid multi cloud to handle the big data issues by improving the performance of cloud management. *International Journal of Computer Applications*, *975*, 8887.

Aslam, K. N., & Swaraj, K. (2019). *Data deduplication with encrypted big data management in cloud computing*. Paper presented at the 2019 International Conference on Communication and Electronics Systems (ICCES).

Ayala, I. D. C. L., Vega, M., & Vargas-Lombardo, M. (2013). Emerging threats, risk and attacks in distributed systems: Cloud computing. In *Innovations and advances in computer, information, systems sciences, and engineering* (Vol. *152*, pp. 37–51): Springer.

Baek, J., Vu, Q. H., Liu, J. K., Huang, X., & Xiang, Y. (2014). A secure cloud computing based framework for big data information management of smart grid. *IEEE Transactions on Cloud Computing*, *3*(2), 233–244.

Bagheri, M., Nurmanova, V., Abedinia, O., Naderi, M. S., Ghadimi, N., & Naderi, M. S. (2018). *Impacts of renewable energy sources by battery forecasting on smart power systems*. Paper presented at the 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe).

Cao, Z., Lin, J., Wan, C., Song, Y., Zhang, Y., & Wang, X. (2016). Optimal cloud computing resource allocation for demand side management in smart grid. *IEEE Transactions on Smart Grid*, *8*(4), 1943–1955.

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Celesti, A., Fazio, M., Romano, A., & Villari, M. (2016). *A hospital cloud-based archival information system for the efficient management of HL7 big data*. Paper presented at the 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), (pp. 406–411).

Chaudhary, R., Aujla, G. S., Kumar, N., & Rodrigues, J. J. (2018). Optimized big data management across multi-cloud data centers: Software-defined-network-based analysis. *IEEE Communications Magazine*, *56*(2), 118–126.

Chaudhuri, S. (2012). What next? A half-dozen data management research goals for big data and the cloud. Paper presented at the Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Systems.

Chen, J., & Dou, H. (2020). *The information strategy of university education and teaching management in the era of cloud computing and big data*. Paper presented at the International conference on Big Data Analytics for Cyber-Physical-Systems.

Chen, D., Gao, H., & Ma, Y. (2021a). Human capital-driven acquisition: Evidence from the inevitable disclosure doctrine. *Management Science*, *67*(8), 4643–4664.

Chen, J., Liu, Y., Xiang, Y., & Sood, K. (2021b). RPPTD: Robust privacy-preserving truth discovery scheme. *IEEE Systems Journal*, *15*, 1–8.

Chen, Y., & Sivakumar, V. (2021). Investigation of finance industry on risk awareness model and digital economic growth. *Annals of Operations Research*, *310*, 1–22.

Cheng, H., Shojafar, M., Alazab, M., Tafazolli, R., & Liu, Y. (2021). PPVF: Privacy-preserving protocol for vehicle feedback in cloud-assisted VANET. *IEEE Transactions on Intelligent Transportation Systems*, *22*, 1–13.

Chiuchisan, I., Costin, H.-N., & Geman, O. (2014). *Adopting the Internet of Things technologies in health care systems*. Paper presented at the 2014 International Conference and Exposition on Electrical and Power Engineering (EPE).

Das, A. K., Adhikary, T., Razzaque, M. A., Alrubaian, M., Hassan, M. M., Uddin, M. Z., & Song, B. (2017). Big media health-care data processing in cloud: A collaborative resource management perspective. *Cluster Computing*, *20*(2), 1599–1614.

Dehghani, M., Ghiasi, M., Niknam, T., Kavousi-Fard, A., Shasadeghi, M., Ghadimi, N., & Taghizadeh-Hesary, F. (2020). Blockchain-based securing of data exchange in a power transmission system considering congestion management and social welfare. *Sustainability*, *13*(1), 1–1.

Dehghani, M., Ghiasi, M., Niknam, T., Kavousi-Fard, A., Shasadeghi, M., Ghadimi, N., & Taghizadeh-Hesary, F. (2021). Blockchain-based securing of data exchange in a power transmission system considering congestion management and social welfare. *Sustainability*, *13*(1), 90.

Esmailiyan, M., Amerizadeh, A., Vahdat, S., Ghodsi, M., Doewes, R. I., & Sundram, Y. (2021). Effect of different types of aerobic exercise on individuals with and without hypertension: An updated systematic review. *Current Problems in Cardiology*, *47*, 101034. doi: https://doi.org/10.1016/j.cpcardiol.2021.101034.

Ganguly, S., Consul, A., Khan, A., Bussone, B., Richards, J., & Miguel, A. (2016). *A practical approach to hard disk failure prediction in cloud platforms: Big data model for failure management in datacenters*. Paper presented at the 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService).

Grander, G., da Silva, L. F., & Gonzalez, E. D. R. S. (2021). Big data as a value generator in decision support systems: A literature review. *Revista de Gestão*, *28*, 205–222.

Gupta, S., & Godavarti, R. (2020). IoT data management using cloud computing and big data technologies. *International Journal of Software Innovation (IJSI)*, *8*(4), 50–58.

Hameurlain, A., & Morvan, F. (2015). Big data management in the cloud: Evolution or crossroad? *Beyond databases, architectures and structures. Advanced technologies for data mining and knowledge discovery* (Vol. *613*, pp. 23–38). BDAS 2015 2016. Cham: Communications in Computer and Information Science, Springer.

Heidari, A., Jabraeil Jamali, M. A., Jafari Navimipour, N., & Akbarpour, S. (2020). Internet of Things offloading: Ongoing issues, opportunities, and future challenges. *International Journal of Communication Systems*, *33*(14), e4474.

Heidari, A., & Navimipour, N. J. (2021). Service discovery mechanisms in cloud computing: A comprehensive and systematic literature review. *Kybernetes*, *51*, 952–981.

Hey, A. J., Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: Data-intensive scientific discovery* (Vol. *1*). Redmond, WA: Microsoft Research.

Hu, T., Wang, S., She, B., Zhang, M., Huang, X., Cui, Y., … Wang, X. (2021). Human mobility data in the COVID-19 pandemic: Characteristics, applications, and challenges. *International Journal of Digital Earth*, *14*, 1126–1147.

Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, *2*, 652–687.

Huang, J., Guo, P., Xie, Q., & Meng, X. (2015). Cloud services platform based on big data analytics and its application in livestock management and marketing. *Proceedings of Science*, *18*, 2015.

Hussien, N. S., & Sulaiman, S. (2016). Mobile cloud computing architecture on data management for big data storage. *International Journal of Advances in Soft Computing and its Applications*, *8*, 139–160.

Inamdar, Z., Raut, R., Narwane, V. S., Gardas, B., Narkhede, B., & Sagnak, M. (2020). A systematic literature review with bibliometric analysis of big data analytics adoption from period 2014 to 2018. *Journal of Enterprise Information Management*, *34*, 101–139.

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Ionescu, L., & Andronie, M. (2021). *Big data management and cloud computing: Financial implications in the digital world*. Paper presented at the SHS Web of Conferences.

Islam, M. T., & Buyya, R. (2019). Resource management and scheduling for big data applications in cloud computing environments. In *Handbook of research on cloud computing and Big data applications in IoT* (pp. 1–23). Australia: IGI Global.

Jain, N. (2020). Secured cloud computing for data management using big data for small and medium educational institutions. *International Journal of Computer Engineering and Technology*, *11*(2), 21–30.

Jamali, J., Bahrami, B., Heidari, A., Allahverdizadeh, P., & Norouzi, F. (2020). Towards the Internet of Things.

Kars-Unluoglu, S., & Kevill, A. (2021). Emotional foundations of capability development: An exploration in the SME context. *Journal of Management & Organization*, *27*, 1–20.

Kaseb, A. S., Mohan, A., & Lu, Y.-H. (2015). *Cloud resource management for image and video analysis of big data from network cameras*. Paper presented at the 2015 International Conference on Cloud Computing and Big Data (CCBD).

Khan, S., Shakil, K. A., & Alam, M. (2016). Cloud-based big data management and analytics for scholarly resources: Current trends, challenges and scope for future research. *arXiv preprint arXiv*:1606.01808.

Lan, Y.-C., & Unhelkar, B. (2015). Sharing big data driven insights using cloud based knowledge management platform: a case study for small and medium enterprises in Taiwan. Paper presented at the Proceedings of the 20th International Conference on Transformative Science and Engineering, Business and Social Innovation, November 1–5, 2015, Fort Worth, Texas.

Lee, J., Lapira, E., Bagheri, B., & Kao, H.-A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing letters*, *1*(1), 38–41.

Li, Y. (2019). *Research on management accounting teaching based on cloud accounting system under big data background*. Paper presented at the 2019 International Conference on Advanced Education, Service and Management.

Li, S. (2021). *Research on the application of cloud accounting in government accounting under the background of big data*. Paper presented at the *Journal of Physics: Conference Series*.

Li, X., Song, J., & Huang, B. (2016). A scientific workflow management system architecture and its scheduling based on cloud service platform for manufacturing big data analytics. *The International Journal of Advanced Manufacturing Technology*, *84*(1–4), 119–131.

Liu, F., Zhang, G., & Lu, J. (2020). Multi-source heterogeneous unsupervised domain adaptation via fuzzy-relation neural networks. *IEEE Transactions on Fuzzy Systems*, *29*, 3308–3322.

Maroli, A., Narwane, V. S., & Gardas, B. B. (2021). Applications of IoT for achieving sustainability in agricultural sector: A comprehensive review. *Journal of Environmental Management*, *298*, 113488.

Munir, R., Wei, Y., Ullah, R., Hussain, I., Arshid, K., & Tariq, U. (2020). Big data of home energy management in cloud computing. *Journal of Quantum Computing*, *2*(4), 193.

Nosratabadi, S., Mosavi, A., Shamshirband, S., Kazimieras Zavadskas, E., Rakotonirainy, A., & Chau, K. W. (2019). Sustainable business models: A review. *Sustainability*, *11*(6), 1663.

Okay, F. Y., & Ozdemir, S. (2016). *A fog computing based smart grid model*. Paper presented at the 2016 international symposium on networks, computers and communications (ISNCC).

Omanović-Mikličanin, E., Maksimović, M., & Vujović, V. (2015). The future of healthcare: Nanomedicine and internet of nano things. *Folia Medica Facultatis Medicinae Universitatis Saraeviensis*, *50*(1), 23–28.

Park, J., Kim, H., Jeong, Y. S., & Lee, E. (2014). Two-phase grouping-based resource management for big data processing in mobile cloud computing. *International Journal of Communication Systems*, *27*(6), 839–851.

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, 1–18.

Pinheiro, E., Weber, W.-D., & Barroso, L. A. (2007). Failure trends in a large disk drive population.

Puthal, D., Nepal, S., Ranjan, R., & Chen, J. (2016). *A secure big data stream analytics framework for disaster management on the cloud*. Paper presented at the 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS).

Roshandeh, A. M., Poormirzaee, R., & Ansari, F. S. (2014). Systematic data management for real-time bridge health monitoring using layered big data and cloud computing. *International Journal of Innovation and Scientific Research*, *2*(1), 29–39.

Sahal, R., Khafagy, M. H., & Omara, F. A. (2016). A survey on SLA management for cloud computing and cloud-hosted big data analytic applications. *International Journal of Database Theory and Application*, *9*(4), 107–118.

Shahid, F., Ashraf, H., Ghani, A., Ghayyur, S. A. K., Shamshirband, S., & Salwana, E. (2020). PSDS – Proficient security over distributed storage: A method for data transmission in cloud. *IEEE Access*, *8*, 118285–118298.

Shamshirband, S., Fathi, M., Chronopoulos, A. T., Montieri, A., Palumbo, F., & Pescapè, A. (2020). Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, and open research issues. *Journal of Information Security and Applications*, *55*, 102582.

Shan, Y.-C., Chao, L., Zhang, Q.-Y., & Tian, X.-Y. (2017). Research on mechanism of early warning of health management based on cloud computing and big data. Paper presented at the Proceedings of the 23rd International Conference on Industrial Engineering and Engineering Management 2016.

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Shen, H., Zhang, M., Wang, H., Guo, F., & Susilo, W. (2021). A cloud-aided privacy-preserving multi-dimensional data comparison protocol. *Information Sciences*, *545*, 739–752.

Simpson, A. V., Farr-Wharton, B., & Reddy, P. (2020). Cultivating organizational compassion in healthcare. *Journal of Management & Organization*, *26*(3), 340–354.

Sinaeepourfard, A., Krogstie, J., & Petersen, S. A. (2018). A big data management architecture for smart cities based on fog-to-cloud data management architecture.

Song, J., Cui, Y., Li, M., Qiu, J., & Buyya, R. (2014). *Energy-traffic tradeoff cooperative offloading for mobile cloud computing*. Paper presented at the 2014 IEEE 22nd International Symposium of Quality of Service (IWQoS) (pp. 284–289).

Sreekanth, R., Rao, G. V. M., & Nanduri, S. (2015). Big data electronic health records data management and analysis on cloud with mongodb: A NoSQL database. *International Journal of Advanced Engineering and Global Technology*, *3*(7), 943–949.

Stergiou, C. L., Psannis, K. E., & Gupta, B. B. (2020). InFeMo: flexible big data management through a federated cloud system. (Vol. 22, pp. 1–22).

Sun, Y. (2021). *Construction and research of digital archives cloud platform based on big data management*. Paper presented at the Journal of Physics: Conference Series.

Suresh, P., Keerthika, P., Sathiyamoorthi, V., Logeswaran, K., Sentamilselvan, K., Sangeetha, M., & Sagana, C. (2021). Cloud-based big data analysis tools and techniques towards sustainable smart city services. In *Decision support systems and industrial IoT in smart grid, factories, and cities* (pp. 63–90). India: IGI Global.

Terrazas, G., Ferry, N., & Ratchev, S. (2019). A cloud-based framework for shop floor big data management and elastic computing analytics. *Computers in Industry*, *109*, 204–214.

Thanigaivasan, V., Narayanan, S. J., Iyengar, S. N., & Ch, N. (2018). Analysis of parallel SVM based classification technique on healthcare using big data management in cloud storage. *Recent Patents on Computer Science*, *11*(3), 169–178.

Ullah, S., Awan, M. D., & Sikander Hayat Khiyal, M. (2018). Big data in cloud computing: A resource management perspective. *Scientific Programming*, *2018*.

Vahdat, S. (2021). The role of IT-based technologies on the management of human resources in the COVID-19 era. *Kybernetes*, ahead-of-print.

Vahdat, S., & Shahidi, S. (2020). D-dimer levels in chronic kidney illness: A comprehensive and systematic literature review. *Proceedings of the National Academy of Sciences, India. Section B: Biological Sciences*, *90*, 1–18.

Waga, D. (2013). Environmental conditions' big data management and cloud computing analytics for sustainable agriculture. SSRN 2349238.

Wang, J., Baker, T., Balazinska, M., Halperin, D., Haynes, B., Howe, B., … Mehta, P. (2017). *The Myria big data management and analytics system and cloud services*. Paper presented at the CIDR.

Wang, K., & Li, S. (2021). Robust distributed modal regression for massive data. *Computational Statistics & Data Analysis*, *160*, 107225.

Wang, L., Ma, Y., Yan, J., Chang, V., & Zomaya, A. Y. (2018). pipsCloud: High performance cloud computing for remote sensing big data management and processing. *Future Generation Computer Systems*, *78*, 353–368.

Wang, K., Wang, H., & Li, S. (2021). Renewable quantile regression for streaming datasets. *Knowledge-Based Systems*, *235*, 107675.

Wang, Z., & Zhao, H. (2016). *Empirical study of using big data for business process improvement at private manufacturing firm in cloud computing*. Paper presented at the 2016 IEEE 3rd international conference on cyber security and cloud computing (CSCloud).

Wang, J., & Zhao, B. (2021). Intelligent system for interactive online education based on cloud big data analytics. *Journal of Intelligent & Fuzzy Systems(Preprint)*, *42*, 1–11.

Xiang, D., Zhang, Y., & Worthington, A. C. (2018). *Determinants of the use of fintech finance among Chinese small and medium-sized enterprises*. Paper presented at the 2018 IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE).

Xiaona, M. (2021). *Informatization strategies of education and teaching management in the era of cloud computing and big data*. Paper presented at the Journal of Physics: Conference Series.

Yan, X., & Nanyun, X. (2020). *Application of cloud accounting in comprehensive budget management of agricultural enterprises under big data*. Paper presented at the E3S Web of Conferences.

Yang, Y. (2018). Research on enterprise cloud accounting and effectiveness management system under big data and internet environment. Paper presented at the Institute of Management Science and Industrial Engineering. Proceedings of 2018 International Workshop on Advances in Social Sciences (IWASS 2018).

Yang, Z., Ghadamyari, M., Khorramdel, H., Alizadeh, S. M. S., Pirouzi, S., Milani, M., … Ghadimi, N. (2021). Robust multi-objective optimal design of islanded hybrid system with renewable and diesel sources/stationary and mobile energy storage systems. *Renewable and Sustainable Energy Reviews*, *148*, 111295.

Yi, H. (2021). Secure social Internet of Things based on post-quantum blockchain. *IEEE Transactions on Network Science and Engineering*, *8*, 1–1.

Zhang, M., Chen, Y., & Susilo, W. (2020). PPO-CPQ: A privacy-preserving optimization of clinical pathway query for e-healthcare systems. *IEEE Internet of Things Journal*, *7*(10), 10660–10672.

https://doi.org/10.1017/jmo.2022.17 Published online by Cambridge University Press

Zhang, B., Fang, B., Yin, J., & Yu, X. (2018). *Research and development of university physical education cloud platform management system based on big data analysis*. Paper presented at the 2018 International Conference on Management, Economics, Education, Arts and Humanities (MEEAH 2018) (Vol. 291, pp. 180–184).

Zhong, L., Fang, Z., Liu, F., Yuan, B., Zhang, G., & Lu, J. (2021). Bridging the theoretical bound and deep algorithms for open set domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, *32*, 1–15.

Zhu, Y., Tan, Y., Luo, X., & He, Z. (2018). Big data management for cloud-enabled geological information services. *Scientific Programming*, *2018*.

Zhuang, M., Zhu, W., Huang, L., & Pan, W.-T. (2021). Research of influence mechanism of corporate social responsibility for smart cities on consumers' purchasing intention. *Library Hi Tech*, *Vol. ahead-of-print*, Vol. ahead-of-print.

Zuo, X. (2017). *Research on enterprise's comprehensive budget management system in the view of big data and cloud accounting*. Paper presented at the 4th International Conference on Education, Management, Arts, Economics and Social Science (ICEMAESS 2017).

Cite this article: Qi W, Sun M, Hosseini SRA (2023). Facilitating big-data management in modern business and organizations using cloud computing: a comprehensive study. *Journal of Management & Organization* **29**, 1697–723. https://doi.org/10.1017/jmo.2022.17