

# A Comprehensive Survey of Big Data Mining Approaches in Cloud Systems

1<sup>st</sup> Zainab Salih Ageed  
Translation Dept.

Nawroz University  
Duhok, Iraq

[zainab.ageed@nawroz.edu.krd](mailto:zainab.ageed@nawroz.edu.krd)

2<sup>nd</sup> Subhi R. M. Zeebaree  
Culture Center

Duhok Polytechnic University  
Duhok, Iraq

[subhi.rafeeq@dpu.edu.krd](mailto:subhi.rafeeq@dpu.edu.krd)

3<sup>rd</sup> Mohammed A. M.Sadeeq  
Quality Assurance

Duhok Polytechnic University  
Duhok, Iraq

[mohammed.abdulrazaq@dpu.edu.krd](mailto:mohammed.abdulrazaq@dpu.edu.krd)

4<sup>th</sup> Shakir Fattah Kak

Dept. Information Technology  
Duhok Polytechnic University  
Duhok -Iraq

[shakir.fattah@dpu.edu.krd](mailto:shakir.fattah@dpu.edu.krd)

5<sup>th</sup> Hazha Saeed Yahia

Information Technology Dept.  
Lebanese French University  
Erbil, Iraq

[hazha.yahya@lfu.edu.krd](mailto:hazha.yahya@lfu.edu.krd)

6<sup>th</sup> Mayyadah R. Mahmood

Computer Science Dept  
University of Zakho  
Duhok, Iraq

[mayyadah.mahmood@uoz.edu.krd](mailto:mayyadah.mahmood@uoz.edu.krd)

7<sup>th</sup> Ibrahim Mahmood Ibrahim  
Information Technology Dept.  
Duhok Polytechnic University  
Duhok -Iraq

[ibrahim.mahmmod@dpu.edu.krd](mailto:ibrahim.mahmmod@dpu.edu.krd)

<https://doi.org/10.48161/qaj.v1n2a46>

**Abstract**—Cloud computing, data mining, and big online data are discussed in this paper as hybridization possibilities. The method of analyzing and visualizing vast volumes of data is known as the visualization of data mining. The effect of computing conventions and algorithms on detailed storage and data communication requirements has been studied. When researching these approaches to data storage in big data, the data analytical viewpoint is often explored. These terminology and aspects have been used to address methodological development as well as problem statements. This will assist in the investigation of computational capacity as well as new knowledge in this area. The patterns of using big data were compared in many articles. In this paper, we research Big Data Mining Approaches in Cloud Systems and address cloud-compatible problems and computing techniques to promote Big Data Mining in Cloud Systems.

**Keywords**— Big Data, Data Computing, Big Data Mining, Cloud Computing

## I. INTRODUCTION

Since cloud computing requires time and has a special design, the researchers turned parallel processing in the cloud ten years ago during the cloud growth. Furthermore, it offers a maximum guaranteed power during processing, and all of its servers are hybrid in terms of computer science [1]. Cloud computing is a form of computing that makes resources available in a dynamically scalable and frequently virtualized manner. It is a form of computing in which resources are distributed as services over the Internet and are dynamically scalable and frequently virtualized [2, 3]. It has also created a medium for viewing and sharing (Big Data Mining) technology, abbreviated as BDMT [4]. These data also contain artificially inserted multi-dimensional information to

avoid privacy violations, posing a serious problem for data mining in the big data sharing process [5]. For incorporating big data sharing and information visualization, cloud computing is a powerful method [6]. Cloud storage is a pay-per-use model that allows networks to access data and demand easily [7]. These services can be easily provided by entering the configurable computing resources shared pool (resources include: networks, servers, storage, applications, and operations), requiring little administration work or communication with service providers, and providing fast and on-demand network access [8, 9]. These services can be distributed easily with minimum administrative effort and contact with service providers [10]. There is one basic concept offered by cloud computing: there is no need to be concerned with how the purchasing services are delivered [11]. You can forget about the complicated fore because the web service takes care of the reliability assessment and can concentrate on your tasks [12]. It depends on consuming the entire regulator for PCs and extensive transparency and cloud storage variants [13]. Developing two basic principles, private and public (i.e. cloud counterpart Intranet and Internet, Online electronic messages and approved facilities similar to those offered by Google, are among the most well-known examples of global one) [14].

The paper structure: have the Cloud Systems in Section II, Big Data Mining In Section III, Literature Review in Section IV, comparison and discussion in section V, Finally, the conclusion in section VI.

## II. CLOUD SYSTEMS

There is one basic concept of cloud computing: there is no need to be concerned with how your buying service is delivered. With web-based services, you can forget about complicated calculations because you will concentrate on the job you are working on [15]. Web-based services rely on consuming the entire regulatory framework for PC organizations and extensive accountability and cloud computing variants [16, 17]. Developing two basic concepts, private and public (i.e. cloud counterparts Intranets besides Internet, Online electronic message plus permitted facilities similar those Google gives which considered utmost acquainted cases of global one) [18]. Client machines, Distributed Servers, and Datacenters are the three basic segments (components) of cloud computing [19]. The components of the cloud are shown in Figure 1.

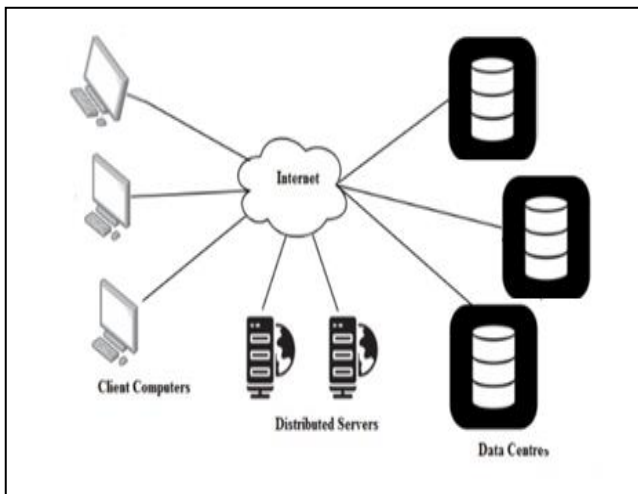


Fig. 1. Cloud Systems [20]

Data exchange in the cloud computers allows multiple users to openly share community data, which increases working efficiency and broadly applicate in cooperative environments [21, 22]. However, data privacy maintenance in a community and the successful exchange of externalized data are great challenges in a group setting [23]. The cloud utilities Today have several providers, including Dropbox, iCloud and Google, that are popular to the general public [24]. Many of these cloud providers provide a variety of services, which maintain files and records on their Internet servers. You will then use any online system to view the same files instead of keeping them on a single screen [25]. This is helpful to have critical data backed up and accessible everywhere in the "internet" in the enterprise definition of daily life. The safety of these services plays a crucial part in developing computing technologies such as the cloud in the security component of the device [26]. After system construction, which is seen by study of software development procedures, security requirements are also taken into consideration [27].

## III. BIG DATA MINING

The Internet, the internets of things, cloud computing, and the widespread use of smart terminals have ushered in the age of big data, characterized by the proliferation of a huge amount of complex and diverse data [28]. Big data is now a strategic tool of tremendous potential value, encouraging industrial upgrading and growth as a key

development factor. It affects scientific reasoning and science methods as well [29, 30]. Although big data has many benefits, such as a large pool of resources and advanced training-calculating technology, it also poses a challenge. As a result of huge, complex, and volatile data, conventional data processing systems are hindered by storage and computational bottlenecks.[31] Gradually improved the working environments of IT professionals and cracked numerous problems for making calculations that include the highest management works, fixing software changes, and using additional calculating series [32]. Big data mining extracts the most valuable information, and expertise from a large, complex, competitive, high-volume, low-density data set and delivers it to the user as a service [33]. It seeks to unearth valuable knowledge and skills, as opposed to traditional data mining. There are, however, gaps in terms of technological history, data environments faced, and scope of mining [34]. The diagram below illustrates the structure of big data based on data mining techniques. As shown in Figure 2 [35], the framework is divided into supporting networks, functional layers, and facilities.

- Platform Layer Support Platforms can build cloud environments with strong and abundant resources. Big data mining can be supported by integrating big mixed data with a range of support processing technologies focused on cloud computing [36]. This cloud environment can not only provide data, hardware, and applications to the rest of the world, but it can also measure moving data, allowing complex data from a variety of sources to be pre-processed, analyzed, and mined more efficiently [37].
- The Functional Layer: Based on the user's demands and needs, this layer will analyze and excavate data [38]. The high efficiency of storages and computing, made available to users as visualizations, data sources, and other technology with high scalability and expandability, is important for science, mining, and other resources.
- Service Level Layers: Via clients, big data mining automatically communicates with service providers and consumers. The effects of mining provide the impetus for pre-processing, analyzing, and mining complex data from various sources [39].

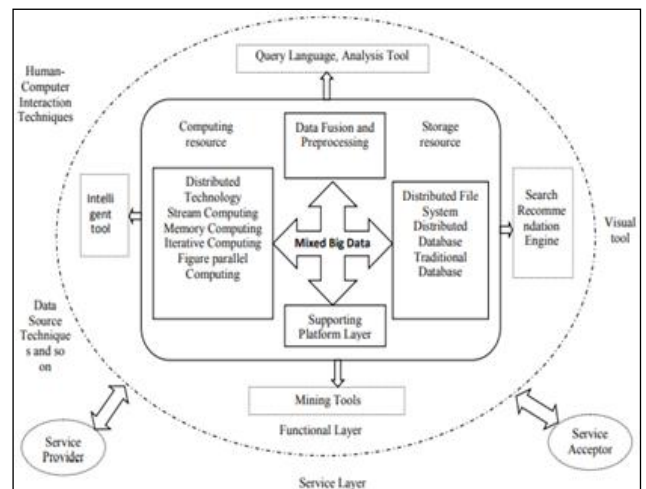


Fig. 2. The Big Data Structure

#### IV. LITERATURE REVIEW

In 2018 Li X [40] introduced the origins, principles, and characteristics of big data's scientific and technical achievement. This paper explores a user habits ontology model and combines it with big data mining technologies to provide customized service schemes for users to suggest scientific and technological accomplishments. For 'login and retrieval,' different logic processes were planned. When a user logs in, the framework uses the user's explicit interest ontology terms as retrieval words to make recommendations, allowing it to gain a higher user recognition level. Users who check in the framework are indicating that they have more in-depth information requirements. This design scheme, based on user behavior ontology models and incorporates explicit desires and implied demands to incorporate customized suggestions, will assist users in accurately gathering knowledge about scientific and technical achievements.

In 2018 Zhu J. et al. [41] looked at WBD principles and data models, wireless data-mining, wireless-knowledge and wireless-knowledges-learning (WKL), and typical practice examples. They aim to encourage and expand WBD research and development opportunities, and such research is useful in developing new theoretical foundations and evolving wireless communications technologies.

In the same year, Ming, J. et al. [42] increased data analysis of mining enterprises based on the characteristics of the techniques and the economics of data mining enterprises. The ANN technique improves the prediction modes of mineral product prices and the prediction and interpolation model of missing geological data. The ANN models created by backpropagation algorithms are used to forecast mineral product prices and predict and interpolate missing geological data. The three-layer NN model was used to establish these prediction and interpolation models.

Wang, Y. et al. in 2018 [43] suggested data security strategies based on standard cloud models. By proposing a dynamic trusts upgrade system to update the direct trust. The dynamic cloud service confidence assessment model, which had been revised, was updated (DCSTEM). The simulation experiment simulates various service levels and offers various cloud services. Five data centers are built in the simulation experiment, each of which corresponds to a different service level.

In 2018 Tomar, D., & Tomar, P. [44] addressed the main features and categorization of Big Data and Hadoop systems that govern the complexity about two case studies used in dealing with Big Data in the cloud. Focus on Cloud Techniques for Big Data and recognize a range of important open research challenges in this field.

In 2018 Suwansrikham, P., & She, K. [45] proposed reducing data privacy threats while also obtaining management access. The large data file is broken down into chunks and spread through various cloud storage providers. Although insider attacks do occur, the attackers only access a portion of the data and cannot reconstruct the entire file. Metadata is generated after the files have been separated. Metadata is where chunk information is stored, such as chunk location, access paths, and data owners' usernames and passwords for connecting to CSP. The research employs the idea of asymmetric defense. The metadata encrypted before being sent to users who request access to the files.

In 2019 Juneja, A., & Das, N. N. [46] suggested a Pre Processing system to enhance the data quality in a weather-monitoring and forecasting application that also takes global warming into account and increases alerts/notifications to warn users and scientists ahead of time. They looked at the Weather Application case study. They tried to develop a system that could forecast weather based on recent global warming issues by integrating Big Data from different sources. However, it emphasizes the importance of approaching data in a Big Data setting early to maximize its utility. This will motivate and prepare companies to take significant steps forward in terms of expansion and strategic plans.

In 2019 Yan, H. et al. [47], the batch gradient descent algorithm based on deep unsupervised learning and the BP algorithm based on Map-Reduce using the parallel neural network model divided by training set, and the CNNRNN model is developed using cloud computing technology. They proposed a cloud-based deep unsupervised learning optimization algorithm was proposed. In contrast to traditional algorithms, the BP algorithm based on Map-Reduce improved neural network training efficiency and demonstrate the feasibility of deep unsupervised learning neural networks using cloud computing.

In 2019 Sudheep, K., & Joseph, S. [48] protected the patient data by proposing a fog computing facility. They demonstrate how to secure big medical data in the cloud using different methods. Healthcare data can be stored and accessed using a decoy technique. They created two photo galleries: one in the cloud layer, the original, and the other in the fog layer.

In 2019 Hu, Y [49] investigated big data with a wide range of potential criminal investigation applications. For intelligent prosecutorial work, the prosecutor's office has begun to use big data. Modern information technologies are used in procuratorate work to create a five-dimensional applications framework of procurator information: "sense, transmission, understanding, usage, and management." "Knowledge" aims to create a procuratorial business knowledge base using cloud computing and big data. "Use" seeks to develop decision-making platforms for procuratorial assistants focused on case-handling data and emerging technologies. Public interest litigation using big data tools can help enhance pollution prevention and regulation, timely rectify unlawful pollutant discharges, provide data support for prosecutorial advice, and quickly repair illegal evidence. The state pays increasing attention to environmental protection.

In 2020 Kotturu, P. K. et al. [50] investigated the probability of data mining hybridization combined with cloud computing. They look at such methods in terms of large data storage from a data analytical perspective. The methodological advancement, as well as the problem statement, has been evaluated based on these aspects. This will assist in the investigation of computational capacity as well as new knowledge in this area.

In 2020 Wang Y. [51] built efficient systems by creating online big data mining and decentralized models. A data connection is another name for association discovery. It establishes association relationships between data entities using the functions of similarity such as Euclidean distances, edit distances, cosine distances, Jaccard functions, and so on.

This parameter is taken into account for the fundamental analysis and taken into consideration in the analysis. The interdependence and inseparability of an enterprise's business processes and internal control structures cannot be overstated. If the organization wants to improve its internal audit and internal control structure, it should improve different aspects of the internal audit institution.

In 2020 Jin L. et al. [52] explained the fundamental concepts of big data and data processing, developed a command information system architecture based on big data analysis. By proposing a solution for real-time data analysis and data delivery for the command information system, and provides a theoretical basis for army command information system in the era of big data.

In 2020 Riddy U et al. [53] classified system traffic attacker using big data-based concentrated log investigation systems using DDOS, SQL Injection, and Bruce Force attacks. Big data was used to detect the attack and square IP addresses. A Big Data-based combined log investigation tool was developed to detect network traffic triggered by aggressors who used DDOS, SQL Injection, and Bruce Force assault. The log record is usually sent to a cloud server, and the investigation begins with many data.

In 2020 Shi Z. et al. [54] proposed a Hadoop-based command and decision system model for effective and accurate ad hoc analysis. They then build a big data research environment for war command and decision-making. With the continued use of cutting-edge technologies in command and control, cloud computing and data mining technology concentrated on big data's characteristics. In the military, the era of big data has arrived. Conventional data processing makes a deep analysis of massive heterogeneous data difficult.

In 2021 Zhao, Jia et al. [55] Proposed System for Online Failure Prediction (Unlabeled Data based online Failure Prediction) UDFP. It implements the analysis-clustering mechanism based on combining the (k-nearest neighbor) KNN and the modelling notion. Analytically, UDFP is shown to alleviate to a degree a supervised learning problem for failure prediction. The experimental findings show that UDFP has prevented manual marking workload, enormous problems as a system solution, increased predictive performance and reduced data processing costs in cloud data centers distributed with security knowledge while improving fault tolerance capacity and strength.

In 2021 Ahmed et al. [56] The method for the study and predictions of COVID-19 was shown for the health monitoring. Big data analysis and IoT are used for the application. The author carries out a descriptive study that focuses on various pandemic signs using a novel real-life dataset, diagnosis, prediction and prescription. This paper's main contribution to diagnosing and predicting a new epidemic consists of combining big data and IoT analytics. The neural network model is developed to help medical personnel detect and forecast the pandemic. In neural networks, the Author expect a pandemic, and they compared the outcomes with other algorithms. The findings show that with a precision rate of 99 per cent, the neural network is doing relatively better.

In 2021 Hongsong et al. [57] suggested deep learning and pipeline machine learning. Security issues in each level of the deep learning system under the Big Data Analysis

Pipeline are thoroughly analyzed. To compare the outcomes of the assault, the Author employ four different attack strategies. The security issues are comprehensively categorized, and the defense responses are analyzed for each security issue. Deployment of drones: The author explores security risks and safety methods in a Drone-Cloud apprenticeship context and chooses Map Engine as a case study. Finally, the potential development of security problems and challenges in the framework of machine learning was proposed.

In 2021 Rouhani et al. [58], a modern approach to collecting and assessing big data systems, was researched and applied to the dominance and inferiority rating. Both functional and non-functional requirements were collected and personalized for the big data platform in the proposed approach. A numerical illustration was given to illustrate the applicability of the method, and in the case of an organization that faced this platform acquisition issue, the recommended solution was used. The proposed solution's key values are the simplified evaluation and selection of the big-data platform and the feasible method. The methodology introduced is useful for Big Data companies, suppliers and sectors, with a view to more precise assessments and improved procurement decisions.

## V. DISCUSSION AND COMPARISON

As seen in Table 1, many studies have been used to explain the research and analyze current trends. It discusses the techniques used, the approaches taken, and the shortcomings that were discovered. The theoretical and explorative approaches and the drawbacks of the mentioned problems are explained in these methods exploration.

We discovered that [40] analyzed user experience modelling and big data mining to provide personalized services. It is discussed how big data analysis can provide knowledge recommendation services for scientific/technological achievement. The construction of user behavior models is proposed, which is combined with customized services in big data environments. [41] Enhanced the data traffic growth in the future as the 5G network grows, as emerging technology like IoT, big data, cloud computing, and artificial intelligence. They want to look at the concept and data model of WBD and WKL, as well as typical practice examples, beginning with the fundamental problems of wireless communication and the interrelationship of demand, environments, and capability, in order to promote and open up further opportunities for WBD research and growth. [42] showed that most missing geological data could be predicted and interpolated, with accurate predictions and interpolation results. The methods for predicting and interpolating technical and economic data are examined. It is discussed and evaluated the regularity of geological data from a single borehole, a group of boreholes, and the regularity of geological data from all boreholes. [43] proposed a model that correctly and efficiently determined the customer population-based on service expectations, increases service requesters' satisfaction, prevents malicious intervention, and is effective and feasible. [44] categorized the big data, an understanding of Hadoop architecture, and a discussion of similar studies concerning big data and cloud platforms. Besides, they discussed the transparent problems of big data in the cloud domain. . [45] concentrated on

sharing massive data files through various cloud storage systems. The data owner separated the file into equal chunks and spread them through various cloud storage providers. They looked at the security situation to ensure security with a system that needs less complexity. [46] analyzed and profiled the big data during the procurement process, then adjust it to requirements to avoid later cost overruns, all while improving and contributing to accurate data analysis. Consequently, even when data is absorbed and used in a big data environment, improving data quality is important. Cleansing to correct as much data as possible, noise filtering to eliminate bad data, and sub-processes for integration and filtering and data transformation/normalization are all being researched. [47] demonstrated the advantages of cloud computing clusters in big data research and the fact that the larger the data, the greater the benefits. It outperforms current algorithms in terms of acceleration ratio, convergence speed, and iteration time. [48] performed a systematic investigation of various strategies for securing big data. Due to the power of user profiling and the decoy strategy, it became apparent during the research that fog computing is capable of delivering good results. [49] constructed a five-dimensional application structure for propuratorial content, modern information technology, namely "sense, transmission, knowledge, use, and management." Prosecutors had begun to hire big data in order to pursue more educated prosecutions. They presented an in-depth study and visualization of different aspects of big data mining. They dive into and expand on the context based on big data exploration and the effects of different analytical methods. [50] discussed the possibilities of hybridizing cloud computing in data mining. Moreover, the analytical data view was discussed in terms of big data storage in exploration of these approaches. The methodological progress and problem statements were evaluated based on

these issues. This helps to explore computing capabilities and new insights into this area. [51] explored the development of a decentralized audit internal control system focused on online big data mining. Finding a way to integrate new technology into internal control is a tempting challenge. The information system is the basis for the IT audit, but it is not a prerequisite. An IT audit would ensure the security of an organization's information system while also offering an objective evaluation of the investment. [52] addressed how command information systems could benefit from big data analysis. How to effectively collect and retrieve unstructured data such as image and video content, and how to handle the systematic analysis of a variety of heterogeneous information data to obtain more useful information knowledge. [53] suggested a technique that is used to address the protection and privacy of massive data sets. They initiate information security focused on big data looking at how big data is secured and how big data can also be used as a digital security weapon. [54] proposed a command and decision system model based on Hadoop to provide effective and accurate ad hoc analysis and a big data analysis environment for war command and decision-making, using cloud storage and data mining technologies. For big data analysis, cloud computing offers the required storage, management, and computing solutions centered on big data's characteristics. [55] Their experimental findings showed that UDFP has prevented manual marking workload, enormous problems as a system solution. [56] focused on Diagnosis and prediction of a new epidemic consist of big combined data and IoT analytics. [57] employed four different attack strategies. The security issues are comprehensively categorized, and the defense responses are analyzed for each security issue. [58] The approach and the proposed solution was used by an enterprise confronted with the platform acquisition problem.

TABLE I. COMPARISON AMONG REVIEWED RESEARCH FOR DATA MINING IN CLOUD SYSTEMS.

Author	Technology and models	Depended Approaches	Achieved Objectives and Significant Results
Li, X. [40] 2018	User activity ontology modelling combined with big data mining technologies	Hadoop is a big data application that uses the MapReduce system. This design scheme, which is based on a user behavior ontology model and incorporates personalized recommendation by combining explicit interest and implied demand, can help users accurately obtain information about scientific and technological achievements in which they are interested.	Provide a scientific, comprehensive, real-time, and reliable big data analysis and decision-making service for personalized intelligence recommendations and intelligent management of scientific and technological achievements.
Zhu J et al. [41] 2018	The data model of WBD	intends to look into the WBD theory and data model, wireless data mining, wireless information, and wireless knowledge learning (WKL), as well as traditional practice examples, to encourage and expand WBD research and development.	For future wireless communications, new theoretical foundations and advanced technologies are being developed. Wireless networking and wireless networks have reached a modern age as a result of new technology. Big wireless data (WBD) is immense, and artificial intelligence (AI) opens up previously unknown possibilities.

Ming, J <b>et al.</b> [42] 2018	A model of artificial neural networks is used to implement the backpropagation algorithm.	The knowledge and value chains within a mining company's operations departments are nearly broken. The company's business decisions cannot keep up with the shifting supply and demand for mineral products. The network has been trained and validated using historical data on price influencing factors for mineral products.	According to the results, the prediction model's practicability is strong, and its prediction accuracy is high.
Wang, Y. <b>et al.</b> [43] 2018	Trust evaluation model	For the conventional cloud model, they have suggested a data protection scheme. They have proposed a framework for real-time updating of confidence. Based on the identification study, it has been used to provide trust. They used a publicly accessible dataset for the experiment. Computational criteria can be extended in terms of analytics and specifications.	For the conventional cloud model, they suggested a data protection scheme. They have proposed a system for real-time confidence updates. It has been used to create trust through recognition analysis. They used a publicly accessible dataset for the experiment. In terms of analytics and requirements, the computational parameters can be expanded.
Tomar, D. <b>et al.</b> [44] 2018	Cloud technology integration between Big Data and Cloud. They used the Hadoop platform and big data. Various factors may be addressed in order to enhance efficiency.	They spoke about how to deploy cloud computing for big data and combine cloud computing with big data. They used the Hadoop platform and big data. Various factors can be addressed in order to enhance efficiency	As cloud computing and big data technology are combined, it creates a powerful combination. They spoke about how to combine big data and cloud computing and how to use cloud computing for big data. They made use of the Hadoop platform and vast volumes of data. To improve performance, a variety of factors can be tackled.
Suwansri kham, P <b>et al.</b> [45] 2018	For big data, an asymmetric safe storage scheme was deployed through multiple cloud providers.	Scheme for asymmetric secure storage They used several cloud providers to implement an asymmetric safe storage scheme for big data. For security research, various computational aspects can be considered.	Scheme for asymmetric secure storage They used several cloud providers to implement an asymmetric safe storage scheme for big data. For security research, various computational aspects can be considered.
Juneja, A. <b>et al.</b> [46] 2019	In a weather monitoring application, pre-processing big data quality is important.	Cleansing to fix as much data as possible, Noise filters to erase bad data, and sub-processes for Integration and Filtering and Data Transformation and Normalization are among the processes being investigated. Data quality is determined by the size, speed, and format in which it is produced. It is impossible to overestimate the importance of high-quality Big Data. We suggest addressing various aspects of the raw data in the pre-processing stage to enhance its accuracy, as the raw data could not be available as-is.	Big Data is evaluated and profiled, and The requirements are then revised to prevent possible cost overruns while also enhancing and resulting in inaccurate data analysis.
Yan, H. et al. [47] 2019	Optimization algorithm for unsupervised learning	They have developed a CNNRNN model. Cloud computing is at the core of it. Backpropagation and the map-reduce algorithm were used. It was employed in the verification and optimization processes.	The performance can be validated using the other approaches.
Sudheep, K. <b>et al.</b> [48] 2019	Medical big data analysis	discussed the different methods used in the health sector. Big data and cloud infrastructure is used to solve this. The anticipated transformation and the applicability gaps have been discussed.	There are no practical implications for permissible sources.

Hu, Y [49] 2019	apply big data investigation technology	Procurators' offices have begun to employ big data in order to do more intelligent work. Procuratorates employ modern information technology to create a five-dimensional application system for procuratorial information: "meaning, delivery, interpretation, usage, and management."	The procuratorate improves case handling performance, maintains continuity in case handling, and reduces resource costs using the big data platform. Creating a big data platform for the procuratorate, with powerful data processing and analysis capabilities, will provide the procuratorate with a clear intellectual guarantee.
Kotturu, P. K. et al. [50] 2020	Data mining hybridization in the context of cloud computing	Study debate, and score the client/server design architecture in terms of the computational ability for data and data chunk adaptation across the entire system.	Addresses the processes, approach, and weaknesses that have been found in the hybridization of data mining. In this methods exploration, the theoretical and explorative approach and the weaknesses in terms of the problems discovered are discussed.
Wang Y. et al. [51] 2020	the decentralized and online big data mining model	The relationship between internal auditing and an organization's internal control structure is interdependent and inseparable. If the company wishes to enhance its internal audit and control structure, it should start by strengthening and developing the internal audit institution as a whole. Build a virtual network card using TUN/TAP technology, then use P2P VPN technology to encrypt data, set up a business gateway, and ensure data protection.	The study proposes a novel intelligent structure for holistic research focused on enterprise management resources in order to create internal control systems that both restrict and enable each other, to enforce and improve the enterprise's internal control system, and to listen to suggestions based on the current state of the enterprise
Jin L et al. [52] 2020	Big data and data analysis builds a framework of the command information system based on big data	It is discussed how big data processing can be used in command information systems. To apply big data analysis to command information systems successfully	Manage the thorough analysis of various heterogeneous information data to obtain more useful information knowledge, efficiently extract and retrieve unstructured battlefield data such as image and video information, and so on.
Reddy, U. et al. [53] 2020	A Big Data-based robust log investigation framework was used to categorize device traffic generated by attackers using DDOS, SQL Injection, and Bruce Force attacks.	Big data was used to determine the assault and square the IP address. A Big Data-based combined log investigation tool was developed to detect device traffic triggered by aggressors who used DDOS, SQL Injection, and Bruce Force attacks. The log record is moved to the centralized cloud server by default, and the big data investigation process begins.	The log file is uploaded to the integrated cloud server, and the big data is used to start the investigation.
Shi Z. et al. [54] 2020	Big data and cloud computing are two technologies that are becoming increasingly popular.	This paper proposes a Hadoop-based command and decision system for providing efficient and precise ad hoc analysis using cloud storage and data mining technologies.	Combining the characteristics of Big Data, cloud computing, and data mining, they provide a big data analysis platform for military command and decision-making.
Zhao, Jia et al. [55] 2021	System for Online Failure Prediction (Unlabeled Data based online Failure Prediction)	It implements the analysis-clustering mechanism based on combining the (k-nearest neighbor) KNN and the modelling notion.	The experimental findings show that UDFP has prevented manual marking workload, enormous problems as a system solution, increased predictive performance and reduced data processing costs in cloud data centers distributed with security knowledge.
Ahmed et al. [56] 2021	The method for the study and predictions of COVID-19 was	Big data analysis and IoT are used for the application. The author carries out a descriptive study that focuses on various	Diagnosis and prediction of a new epidemic consist of big combined data and IoT analytics. The neural network

	shown for the health monitoring	pandemic signs using a novel real-life dataset, diagnosis, prediction and prescription.	model is developed to help medical personnel detect and forecast the pandemic.
Hongson g et al. [57] 2021	suggested deep learning and pipeline machine learning.	Security issues in each level of the deep learning system under the Big Data Analysis Pipeline are thoroughly analyzed	employ four different attack strategies. The security issues are comprehensively categorized, and the defense responses are analyzed for each security issue.
Rouhani et al. [58] 2021	A modern approach to the collection and assessment of big data systems was researched and applied to the dominance and inferiority rating	Both functional and non-functional requirements were collected and personalized for the big data platform in the proposed approach.	The use of the approach and the proposed solution was used by an enterprise confronted with the platform acquisition problem. The key values of the solution proposed are the simplified evaluation and selection of the big data.

## VI. CONCLUSION

The amount of data available constrains the research and decision-making process. Different aspects of data mining in the context of Big Data and Big Data Mining are discussed in this paper, examined and visualized in great detail. This paper discussed how big data analysis could be used to provide knowledge recommendation services. Besides, investigated the context in which big data research takes place to provide personalized services, the results of various analytical approaches, and a discussion of similar studies concerning big data and cloud platforms. It illustrated the methodology's integration in the direction of big data and cloud analytics for data access and analysis alignment mechanisms by illustrating the full generalization of fifteen empirical and methodological studies. These studies looked at the security situation to ensure security with a system that needs less complexity-improving, contributing to accurate data analysis and demonstrating the advantages of cloud computing clusters in big data research. The larger the data, the greater the benefits.

## REFERENCES

- [1] Z. Ageed, M. R. Mahmood, M. Sadeeq, M. B. Abdulrazzaq, and H. Dino, "Cloud computing resources impacts on heavy-load parallel processing approaches," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 22, pp. 30-41, 2020.
- [2] H. H. Song, "Testing and evaluation system for cloud computing information security products," *Procedia Computer Science*, vol. 166, pp. 84-87, 2020.
- [3] O. Alzakholi, H. Shukur, R. Zebari, S. Abas, and M. Sadeeq, "Comparison among cloud technologies and cloud performance," *Journal of Applied Science and Technology Trends*, vol. 1, pp. 40-47, 2020.
- [4] Y. S. Jghef and S. Zeebaree, "State of Art Survey for Significant Relations between Cloud Computing and Distributed Computing," *International Journal of Science and Business*, vol. 4, pp. 53-61, 2020.
- [5] H. Li, T. Wang, and J. Jiang, "Exploration on the Ideological and Political Development Dynamics Investigation and Judgment of College Students Considering Big Data Mining Technology," in *2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE)*, 2018, pp. 740-745.
- [6] Y. Sun and M. Liu, "On the dynamic integration of socialist core values and ideological and political education of college students," in *2014 International Conference on Education Technology and Social Science*, 2014, pp. 179-182.
- [7] M. M. Sadeeq, N. M. Abdulkareem, S. R. Zeebaree, D. M. Ahmed, A. S. Sami, and R. R. Zebari, "IoT and Cloud Computing Issues, Challenges and Opportunities: A Review," *Qubahan Academic Journal*, vol. 1, pp. 1-7, 2021.
- [8] L. Ruixue and R. Youqun, "Research on flow experience and empathy effect in immersive virtual environment," *[J], Electrification Education Research*, vol. 40, pp. 99-105, 2019.
- [9] D. A. Hasan, B. K. Hussan, S. R. Zeebaree, D. M. Ahmed, O. S. Kareem, and M. A. Sadeeq, "The Impact of Test Case Generation Methods on the Software Performance: A Review," *International Journal of Science and Business*, vol. 5, pp. 33-44, 2021.
- [10] C.-C. Chen, K.-W. Lee, C.-C. Chang, D.-N. Yang, and M.-S. Chen, "Efficient large graph pattern mining for big data in the cloud," in *2013 IEEE International Conference on Big Data*, 2013, pp. 531-536.
- [11] D. H. Maulud, S. R. Zeebaree, K. Jacksi, M. A. M. Sadeeq, and K. H. Sharif, "State of Art for Semantic Analysis of Natural Language Processing," *Qubahan Academic Journal*, vol. 1, 2021.
- [12] Z. N. Rashid, S. R. Zeebaree, and A. Shengul, "Design and analysis of proposed remote controlling distributed parallel computing system over the cloud," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 118-123.
- [13] S. R. Zeebaree, "Remote Controlling Distributed Parallel Computing System over the Cloud (RCDPCSC)," in *2020 3rd International Conference on Engineering Technology and its Applications (IICETA)*, 2020, pp. 258-258.



- [14] H. Shukur, S. R. Zeebaree, A. J. Ahmed, R. R. Zebari, O. Ahmed, B. S. A. Tahir, *et al.*, "A State of Art Survey for Concurrent Computation and Clustering of Parallel Computing for Distributed Systems," *Journal of Applied Science and Technology Trends*, vol. 1, pp. 148-154, 2020.
- [15] Z. N. Rashid, S. R. Zebari, K. H. Sharif, and K. Jacksi, "Distributed cloud computing and distributed parallel computing: A review," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 167-172.
- [16] L. M. Haji, S. Zeebaree, O. M. Ahmed, A. B. Sallow, K. Jacksi, and R. R. Zebari, "Dynamic resource allocation for distributed systems and cloud computing," *TEST Engineering & Management*, vol. 83, pp. 22417-22426, 2020.
- [17] A. B. Sallow, H. I. Dino, Z. S. Ageed, M. R. Mahmood, and M. B. Abdulrazaq, "Client/Server Remote Control Administration System: Design and Implementation," *Int. J. Multidiscip. Res. Publ.*, vol. 3, p. 7, 2020.
- [18] K. Jacksi, S. Zeebaree, and N. Dimililer, "Design and Implementation of LOD Explorer: A LOD Exploration and Visualization Model," *Journal of Applied Science and Technology Trends*, vol. 1, pp. 31-39, 2020.
- [19] P. Y. Abdullah, S. R. Zeebaree, H. M. Shukur, and K. Jacksi, "HRM system using cloud computing for Small and Medium Enterprises (SMEs)," *Technology Reports of Kansai University*, vol. 62, p. 04, 2020.
- [20] D. C. Marinescu, *Cloud computing: theory and practice*: Morgan Kaufmann, 2017.
- [21] J. Shen, T. Zhou, D. He, Y. Zhang, X. Sun, and Y. Xiang, "Block design-based key agreement for group data sharing in cloud computing," *IEEE Transactions on Dependable and Secure Computing*, vol. 16, pp. 996-1010, 2017.
- [22] P. Y. Abdullah, S. R. Zeebaree, K. Jacksi, and R. R. Zebari, "An hrms system for small and medium enterprises (sme) s based on cloud computing technology," *International Journal of Research-GRANTHAALAYAH*, vol. 8, pp. 56-64, 2020.
- [23] L. Chen, X. Sun, Z. Xia, and Q. Liu, "An efficient and privacy-preserving semantic multi-keyword ranked search over encrypted cloud data," *International Journal of Security and Its Applications*, vol. 8, pp. 323-332, 2014.
- [24] Z. N. Rashid, S. R. Zeebaree, and A. Sengur, "Novel Remote Parallel Processing Code-Breaker System via Cloud Computing."
- [25] L. M. Haji, O. M. Ahmad, S. R. Zeebaree, H. I. Dino, R. R. Zebari, and H. M. Shukur, "Impact of cloud computing and internet of things on the future internet," *Technology Reports of Kansai University*, vol. 62, pp. 2179-2190, 2020.
- [26] C. M. Mohammed and S. R. Zeebaree, "Sufficient Comparison Among Cloud Computing Services: IaaS, PaaS, and SaaS: A Review," *International Journal of Science and Business*, vol. 5, pp. 17-30, 2021.
- [27] F. E. F. Samann, S. R. Zeebaree, and S. Askar, "IoT Provisioning QoS based on Cloud and Fog Computing," *Journal of Applied Science and Technology Trends*, vol. 2, pp. 29-40, 2021.
- [28] A. I. Abdulla, A. S. Abdurraheem, A. A. Salih, M. A. Sadeeq, A. J. Ahmed, B. M. Ferzor, *et al.*, "Internet of Things and Smart Home Security," *Technol. Rep. Kansai Univ.*, vol. 62, pp. 2465-2476, 2020.
- [29] A. S. Abdurraheem, A. A. Salih, A. I. Abdulla, M. A. Sadeeq, N. O. Salim, H. Abdullah, *et al.*, "Home automation system based on IoT," 2020.
- [30] M. A. Sadeeq, S. R. Zeebaree, R. Qashi, S. H. Ahmed, and K. Jacksi, "Internet of Things security: a survey," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 162-166.
- [31] Z. H. Deng, B. Fan, Y. J. Lu, and Z. F. Li, "Discussion about big data mining based on hadoop," in *Applied Mechanics and Materials*, 2013, pp. 2063-2066.
- [32] Z. S. Ageed, R. K. Ibrahim, and M. A. Sadeeq, "Unified Ontology Implementation of Cloud Computing for Distributed Systems," *Current Journal of Applied Science and Technology*, pp. 82-97, 2020.
- [33] S. S. R. Zeebaree, S. Ameen, and M. Sadeeq, "Social Media Networks Security Threats, Risks and Recommendation: A Case Study in the Kurdistan Region," *International Journal of Innovation, Creativity and Change*, vol. 13, pp. 349-365, 2020.
- [34] C. Stergiou, K. E. Psannis, B.-G. Kim, and B. Gupta, "Secure integration of IoT and cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 964-975, 2018.
- [35] Y. Jiugen and X. Ruonan, "Cloud Computing-based Big Data mining Connotation and Solution," in *2020 15th International Conference on Computer Science & Education (ICCSE)*, 2020, pp. 245-248.
- [36] A. A. Yazdeen, S. R. Zeebaree, M. M. Sadeeq, S. F. Kak, O. M. Ahmed, and R. R. Zebari, "FPGA Implementations for Data Encryption and Decryption via Concurrent and Parallel Computation: A Review," *Qubahan Academic Journal*, vol. 1, pp. 8-16, 2021.
- [37] Z. M. Khalid and S. R. Zeebaree, "Big Data Analysis for Data Visualization: A Review," *International Journal of Science and Business*, vol. 5, pp. 64-75, 2021.
- [38] M. Sadeeq, A. I. Abdulla, A. S. Abdurraheem, and Z. S. Ageed, "Impact of Electronic Commerce on Enterprise Business," *Technol. Rep. Kansai Univ.*, vol. 62, pp. 2365-2378, 2020.
- [39] A. B. Sallow, S. R. Zeebaree, R. R. Zebari, M. R. Mahmood, M. B. Abdulrazaq, and M. A. Sadeeq, "Vaccine Tracker/SMS Reminder System: Design and Implementation."
- [40] X. Li, "Study on information recommendation of scientific and technological achievements based on user behavior modeling and big data mining,"

- in *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 2018, pp. 228-232.
- [41] J. Zhu, C. Gong, S. Zhang, M. Zhao, and W. Zhou, "Foundation study on wireless big data: Concept, mining, learning and practices," *China Communications*, vol. 15, pp. 1-15, 2018.
- [42] J. Ming, L. Zhang, J. Sun, and Y. Zhang, "Analysis models of technical and economic data of mining enterprises based on big data analysis," in *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 2018, pp. 224-227.
- [43] Y. Wang, J. Wen, W. Zhou, and F. Luo, "A novel dynamic cloud service trust evaluation model in cloud computing," in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 2018, pp. 10-15.
- [44] D. Tomar and P. Tomar, "Integration of Cloud Computing and Big Data Technology for Smart Generation," in *2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2018, pp. 1-6.
- [45] P. Suwansrikkham and K. She, "Asymmetric secure storage scheme for big data on multiple cloud providers," in *2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing,(HPSC) and IEEE International Conference on Intelligent Data and Security (IDS)*, 2018, pp. 121-125.
- [46] A. Juneja and N. N. Das, "Big data quality framework: pre-processing data in weather monitoring application," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 559-563.
- [47] H. Yan, P. Yu, and D. Long, "Study on deep unsupervised learning optimization algorithm based on cloud computing," in *2019 international conference on intelligent transportation, Big data & smart city (ICITBS)*, 2019, pp. 679-681.
- [48] K. Sudheep and S. Joseph, "Review on Securing Medical Big Data in Healthcare Cloud," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 2019, pp. 212-215.
- [49] Y. Hu, "Intelligent Procuratorate Depends on Big Data Investigation Technology," in *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 2019, pp. 20-23.
- [50] P. K. Kotturu and A. Kumar, "Data mining visualization with the impact of nature inspired algorithms in big data," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 2020, pp. 664-668.
- [51] Y. Wang, S. Guo, J. Wu, and H. H. Wang, "Construction of Audit Internal Control System Based on Online Big Data Mining and Decentralized Model," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 2020, pp. 623-626.
- [52] L. Jin, M. Xing, and R. Wang, "Operation Framework of the Command Information System Based on Big Data Analysis," in *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, 2020, pp. 459-462.
- [53] E. Brumancia, R. Gomathi, and K. Indira, "Integrative Approach Of Big Data And Network Attacks Analysis In Cloud Environment," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 2020, pp. 314-317.
- [54] Z. Shi, G. Zhao, and J. Liu, "Research on the Model of Command and Decision System for Big Data," in *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, 2020, pp. 481-484.
- [55] J. Zhao, Y. Ding, Y. Zhai, Y. Jiang, Y. Zhai, and M. Hu, "Explore unlabeled big data learning to online failure prediction in safety-aware cloud environment," *Journal of Parallel and Distributed Computing*, 2021.
- [56] I. Ahmed, M. Ahmad, G. Jeon, and F. Piccialli, "A Framework for Pandemic Prediction Using Big Data Analytics," *Big Data Research*, vol. 25, p. 100190, 2021.
- [57] C. Hongsong, Z. Yongpeng, C. Yongrui, and B. Bhargava, "Security Threats and Defensive Approaches in Machine Learning System Under Big Data Environment," *Wireless Personal Communications*, vol. 117, pp. 3505-3525, 2021.
- [58] S. Rouhani and S. Rotbei, "Big data platforms: in the lens of selection and evaluation approach," *Journal of Decision Systems*, pp. 1-30, 2021.