IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 25, 2021, accepted July 21, 2021, date of publication July 26, 2021, date of current version August 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100287

Analyzing and Evaluating Critical Challenges and Practices for Software Vendor Organizations to Secure Big Data on Cloud Computing: An AHP-Based Systematic Approach

ABUDUL WAHID KHAN[1], MASEEH ULLAH KHAN[1], JAVED ALI KHAN[2], ARSHAD AHMAD[3,4], KHALIL KHAN[3], MUHAMMAD ZAMIR[5], WONJOON KIM[6], AND MUHAMMAD FAZAL IJAZ[7], (Member, IEEE)

[1]Department of Computer Science, University of Science and Technology Bannu, Bannu 28100, Pakistan
[2]Department of Software Engineering, University of Science and Technology Bannu, Bannu 28100, Pakistan
[3]Department of IT and Computer Science, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur 22620, Pakistan
[4]Sino-Pak Center for Artificial Intelligence (SPCAI), Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur 22620, Pakistan
[5]Department of Mathematics, University of Science and Technology Bannu, Bannu 28100, Pakistan
[6]Division of Future Convergence (HCI Science Major), Dongduk Women's University, Seoul 02748, Republic of Korea
[7]Department of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, Republic of Korea

Corresponding authors: Wonjoon Kim (wjkim@dongduk.ac.kr) and Muhammad Fazal Ijaz (fazal@sejong.ac.kr)

This work was supported by the Dongduk Women's University Grant No. 2021-05573.

ABSTRACT Recently, its becomes easy to track down the data due to its availability in a large number. Although for data management, processing, and obtainability, cloud computing is considered a well-known approach for organizational development on the internet. Despite many advantages, cloud computing has still numerous security challenges that can affect the big-data usage on cloud computing. To find the security issues/challenges that are faced by software vendors' organizations we conducted a systematic literature review (SLR) through which we have find out 103 relevant research publications by developing a search string that is inspired by the research questions. This relevant data was comprised from different databases e.g. Google Scholar, IEEE Explore, ScienceDirect, ACM Digital Library, and SpringerLink. Furthermore, for the detailed literature review, we have accomplished all the steps in SLR, for example, development of SLR protocol, Initials and final assortment of the relevant data, data extraction, data quality assessment, and data synthesis. We identified fifteen (15) critical security challenges which are: data secrecy, geographical data location, unauthorized data access, lack of control, lack of data management, network-level issues, data integrity, data recovery, lack of trust, data sharing, data availability, asset issues, legal amenabilities, lack of quality, and lack of consistency. Furthermore, sixty four (64) standard practices are identified for these critical security challenges using the proposed SLR that could help vendor organizations to overcome the security challenges for big data. The findings of our research study demonstrate the resemblances and divergences in the identified security challenges in different periods, continents, databases, and methods. The proposed SLR will also support software vendor organizations for securing big data on the cloud computing platforms. This paper has the following content: in Section II, we have describe the Literature review; in Section III, research methodology is specified; in Section IV, the findings of the SLR and the analysis of result are discussed; in Section V, the limitations of this research are given; in Section VI, we discussed our conclusions and future work.

INDEX TERMS Security challenges, big data, cloud computing, SLR, vendor, SPSS.

I. INTRODUCTION

From a decade data is generated exponentially and it can be accessed very easily with effective cost by leaps and bounds [1]. The data volume may be increased from TB to PB because 2.5 quintillion bytes of data may be gathered per day, according to Walmart, where they group and store about 2.5 PB of data in a hour [2]–[4]. This data may be comprised of images, videos, audios, text, social links, GPS signals, commercial, and social accounts, etc. and might be gathered

The associate editor coordinating the review of this manuscript and approving it for publication was M. Anwar Hossain.

from different sources, for example, industries, banks, educational institutes, health departments, and other business organizations, etc. [5], [6]. The data is either in the form of structured, semi-structured, or un-structured. The IDC's (International Data Corporation) verified that the ratio of structured data on internet is around 32% while un-structured data is around 63% [7]. This data in any form can result in big data whenever its volume, velocity, variety, variability, value, visualization, or veracity surpasses the volume of IT system for the storage, processing, and management of that data [3], [8], [9]. Furthermore, whenever the volume of data increased from Terabytes to Zettabytes then it results in big data [10], [77], because at that time massive volume of data will overflow the processing and storing capacity of the IT system [3], [8], [11]–[13], [81].

According to Apache Hadoop, big data is the data that have a large volume and size on which traditional IT system cannot operate, as storing and processing with the time and possessions given [2]. Though big data has no definite size because of its increasing volume from petabytes to exabytes still many organizations have collected this huge volume of data by conducting reviews and surveys all over the world to advance their procedure of decision-making to maintain a good future for their organizational business. Therefore, cloud comes for processing, storing and managing this large volume of data on internet instead of traditional IT systems [11], [12], [14], [66], [68], [69], [80], [81].

To manage this large volume of data for software vendors organizations, cloud computing are used to perform all these processes on the internet instead of local server. Cloud computing is the technology which provides a series of IT related services to their clients over the internet to fulfill their day-today needs. While the quality of these services depends on the delivery of cloud computing possessions, software's, and data available on the internet instead of traditional systems or other devices on user needs with effective costs [8], [9], [11], [15]–[20], [64], [66], [80].

To process and manage this large amount of data, the network of cloud computing is connected to a large number of servers in a connection with each other's and then processed data of the end-user are stored on a secret storing address located any-where in the world [20]–[22], [64], [70]. Cloud computing basically provides three types of services which are ''Software as a Service (SaaS)'', ''Platform as a Service (PaaS)'', and ''Infrastructure as a Service (PaaS)''. Some popular examples of cloud computing are: YouTube, Gmail, LinkedIn, ToTok, Facebook, Twitter, Zoom App, and Dropbox etc. Cloud Computing provides agility, easiness, pliability, and scalability that's why cloud computing is used in organizations for data processing and management very rapidly [1], [8], [12], [17]–[19], [23]–[25], [69], [77].

''Software as a Service (SaaS)'' is a service of cloud computing that carry software application on the internet and does not requires to install the software on the user computer, as they can easily and simply access the services of cloud via the internet, for example Google Workspace, Dropbox, and Salesforce. etc. [1], [13], [22], [23], [25]. ''Platform as a Service (PaaS)'' is a service of cloud computing, which provide a development environment to the users, on which the customers can upload their own software application and coding like ''Windows Azure'', ''Heroku'', ''Force.com'', ''Google App Engine'', etc [1], [13], [22], [23], [25]. While ''Infrastructure as a Service (IaaS)'' is a service of ''cloud computing'' where multiple ''computing resources'' are given to the customers by the IaaS. IaaS services are accessed by the cloud customers by using a wide area network such as internet. IaaS services consist of storage, network, operating system, hardware, and storage devices on user request, for example ''Amazon Web Services (AWS), Microsoft Azure, and Google Compute Engine (GCE)'' [1], [13], [22], [23], [25].

Cloud computing mainly has three types, such as ''public cloud, private cloud, and hybrid cloud. The Public cloud is the computing services provided by the third party providers publicly on the internet. This type of cloud services are available for all kind of users and can be use freely but need to pay for the consumed services of the cloud. The Private cloud is a computing service, which is provided on the internet for the selected customers not for all common users on the internet. While the private cloud provides maximum security and privacy through the internal hosting and firewalls. The Hybrid cloud is the combination of public and private clouds. In this type of cloud computing each cloud is managed independently, however the application and data could be shared between the clouds [23], [10], [26], [62].

Presently cloud computing has become very much useful and popular for storage and processing of the data. It has many advantages like multi-tenancy, resource sharing, data storing and clear virtualization, however cloud computing still faces certain security challenges, which are [1], [16], [17], [22], [59], [63], [67]:

- Confidentiality
- Privacy
- Data Integrity issue
- Data Availability issue
- Trusted Third Party
- Interoperability Issue
- Malicious Insider attack
- Lack of Trust issue
- Losing Control over Data

To address these security challenges we have formulated below two research questions:

RQ1: ''What are the main security issues / challenges faced by the vendor organization to secure the Big Data on Cloud Computing?''

RQ2: ''What are the existed solutions / practices as defined in the literature used for avoiding the security issues / challenges faced by the software vendor organization?''

Keeping in view the above research questions, we find out 15 critical security challenges that might affect big data security on the cloud computing and identified 64 practices that would help in overcoming these security challenges from

the selected 103 research articles shortlisted for the proposed SLR study. The results validates the semblances and differences in the identified security challenges in different periods, continents, databases, and methods. The SLR challenges and practices identification will also assist software vendor's companies to secure big data on the cloud computing. For the proposed literature study, we mainly targeted software vendor organizations that develop software applications using agile software methodology.

II. BACKGROUND

Nowadays in the digital world, the data produced by every organization increasing exponentially and it is hard to be managed by warehouse technology. The large volume of raw data produced by various data sources, required the big data techniques to analyze it [1], [27], [28]. According to Wal-Mart, they processed more than million of user's records with in an hour and store up-to 2.5 petabytes of user's data [3], [4], [29], [30]. Congress library reports that they gathered 235 terabytes of novel data every year and stored 60 petabytes of this data. In 2014, more than 5.5 billion cell phones were used, and each of them can creates terabytes of call records every year [28]. During 2000s, the International Data Corporation (IDC), an international leading market firm reports that that digital world which was 4.4 ZB in the year of 2003, will be grown up-to 44 ZB by 2020 [28], [31].

For big data analytics maximum number of raw data is unstructured data, which is obtained from multiple data sources and application such as weblogs, Facebook, Twitter, LinkedIn, text files, dropbox, emails, images, audio, or video recordings. Big data is destined for handling and managing unstructured data with the help of key value pairs. Big data concept is defined by Gartner and Will Dailey [28], [32], [35]. According to Gartner [28], [32] the big data is that data which has maximum volume, velocity, and variety of information that requires efficient costs, advanced ways of data processing and decision making. Similarly, Will Dailey [28] defined big data as, ''The super processing environment that is engineered for the parallel computing process through huge volume of distributed data so to analyze that data ''.

The large volume of data results in big data, when its volume increases from terabytes to zettabytes [10] and the volume, value, velocity, variety, veracity, variability, and its visibility run-off the storage and processing ability of IT System [3], [8], [11], [12], [35], [70]. Cloud computing has mainly faced security challenges on four levels which are network level, data level, user authentication level, and generic level [21], [59]–[62]. The above mentioned security challenges faced by the big data on the cloud are discussed in detail below:

Distributed Nodes: this is a network level issue where all the nodes are connected with each other and data flow can be occurred any-where across the nodes. User cannot know about the exact node of computation, data flow, and data location. As the data of each user is distributed around the world [15], [21], [33], [34].

Un-encryption of data: For better performance and efficacy cloud computing like Hadoop and MapReduce store and process the data without encryption, due to which critical data of end user are unsecure and malicious insiders can easily get access to the critical data [12], [13], [15], [18], [33], [35]–[37].

Data Recovery issue: whenever in cloud computing a data loss is occurred due to any cause then recovering the complete data is not possible as some of its parts are deleted permanently. It is also considered as a big challenge for cloud computing to find out the exact node of data deletion for recovery in the cloud [13], [16], [26], [38], [39].

Trust issue: Cloud computing has a trust issue because the storage location of data is not the same as that of customer. Moreover, the customer has no idea of their critical data and its security, that's why customer do not trust on cloud computing and refer to store their personal data on their own system at home instead of cloud environment [7], [11], [24], [33], [40], [41].

Technical Network issue: In cloud computing data communication occur through unsecured network such as Internet, where every node can easily change data or inter node communication so that to breach the entire connection of the system [15], [40], [42], [43].

Un-controlled access of Data: In cloud computing all the nodes are inter linked with each other for data processing. As every node has free access to the user data, which can make the possibility for some harmful nodes to change or steal the critical data. So this un-controlled administrative access of each node to the data is a critical issue in cloud computing [13], [15], [10], [36], [38].

Cloud computing Security association also discussed the main challenges of security of big data on cloud environment as [26], [36], [44], [45], [61], [67], [76]:

- Secure the computation in distributed programming frameworks
- Security best practices for non-relational databases
- Secure data storage and transactions logs
- End-point input validation/filtering
- Real-time security monitoring
- Scalable and composable privacy-preserving data mining and analytics
- Cryptographically enforced data centric security
- Granular access control
- Granular audits
- Source of data

It is contended that avoiding, addressing, or justifying these security challenges by software vendors organization can yield in a successful consequences of the big data security on cloud environment. This motivated us, to explore in detailed the security challenges confronted by the software vendors organizations by conducting a Systematic Literature Review (SLR) that helps in identifying critical security challenges faced by big data over the platform of cloud computing. Furthermore, we are more interested in finding

the best solutions or practices, which needs to be executed by software vendor's organization to pointout, avoid, or mitigate these security challenges.

We contributed in certain aspects, Firstly, we have find out through SLR that there are many security challenges that big data faces on the platform of cloud computing. These challenges were discussed by many researchers where they have defined each security challenge separately. To our knowledge, we haven't found any such model using SLR in the literature that supports software vendor organizations on the usage of big data over the platform of cloud computing. Secondly, our security model for Big Data usage on Cloud Computing (SMBDCC) will provide solution for security related challenges. Our security model is a unique contribution to support software vendor organization about security related challenges of big data on cloud computing. Our study is based on, to identify the security challenges by using systematic literature review (SLR) and empirical study (ES), and also to examine the solutions/practices, through SLR and empirical study in the software organizations, so that to address, mitigate, and avoid these security challenges.

III. METHODOLOGGY

We the method of Systematic Literature Review (SLR) approach for the identification of security challenges/issues that will assist software vendor organization for big data usage on cloud computing [46], [24]. According to Gangawane and Devi [24] SLR is a novel method of recognition, analyzation, and gathering of all possible information about a unique technology to know about the novel track and the investigation of research questions. In SLR, the searching of related published work is done with the help of pre-defined search string that is based on the research questions. The SLR used a pre-defined inclusion/exclusion method/criteria for the analyses of the collected data. The SLR consist of three phases, such as, planning a review, conducting a review, and reporting a review [46], [71], [78]. Where, the reliability of the SLR results is greater than ordinary literature review because SLR followed a method of systematic evaluation. For the proposed SLR, we followed step by step procedure to identify the challenges and practices of big data on cloud computing. Where, we select each article based on its relevancy to the topic or search string after that inclusion/exclusion criteria were applied to exclude the irrelevant papers from our search. Next, the data extraction is performed from the selected research papers systematically and then data synthesis & quality of publication is conducted. From the proposed SLR, we initially find out 19 security challenges, which are further reduced to 15 challenges by merging the similar critical security challenges for big data security on cloud.

In our SLR, the planning and conducting phases are already accomplished, and now we need to report the results of conducted phase. The central objective of this research, paper is to highlight all the security challenges/issues and it's appropriate management with the help of SLR.

We have find out total of 15 challenges (as shown in Table 2), which are very much critical challenges for big data usage on cloud computing, These are 'Data secrecy issue', 'Geographical data location issue', 'Unauthorized data access issue', 'Lack of Control', 'Lack of Data Management', 'Network level issues', 'Data integrity issues', 'Data Recovery issues', 'Lack of Trust', 'Data Sharing Issue', 'Data Availability', 'Assets Issue', and 'Legal Amenabilities' on the basis of frequency >=25%, where the same approach used by other researchers [47]. We have used this criteria to consider more challenges to assist the vendor organizations.

A. SEARCH PROCESS AND PRACTICE

To make the search string for the SLR, we have follow the following steps

- Identify population, outcomes, and intervention for the definition of search term.
- Identification of spellings and synonyms of the substitutes.
- Keywords verification and validation for search terms in all related selected literature.
- For accurate searching result use of Boolean operators (AND/OR), if there is any need to guide the search engine.

Firstly, we designed a search string for trial bases on different databases to identify the relevant research articles. This trial search was searched on the five different search engines i.e. Google scholar, ACM, IEEE, Springer link, and Science direct. Below is the trial search string which shows different results on each database but still the results were not satisfactory.

((''Big data'') AND (''Cloud computing'' OR ''Cloud environment'') AND (Challenges OR Barriers OR Issues OR Problems)).

The final search string is:

((vendors OR merchants OR retailers OR contractor OR suppliers) AND (''big data'' OR ''massive data'' OR ''data science'' OR ''data analytics'') AND (''cloud computing'' OR ''cloud environment'' OR ''cloud technology'' OR ''cloud airframe'' OR ''cloud database'') AND (''security challenges'' OR ''security issues'' OR ''security risks'' OR barriers OR ''security problems'') AND (''security practices'' OR ''security reviews'' OR ''security methods'' OR approaches OR procedures OR ''security solutions'')).

The search results of relevant publications that are obtained with the final search string are listed below in TABLE 1.

For the final selection of research articles, as shown in Table. 1, we used the inclusion and exclusion criteria's, where we need to select the relevant research publications based on complete reading, paper quality, and verification. Each of them are elaborated below:

TABLE 1. Data sources and search strategy.

Digital libraries	Total identified publication	Primary Selection	Final Selection
Google Scholar	13,500	430	65
IEEE	41	33	11
ACM	10,25	41	4
Science Direct	7,309	42	1
Springer Link	13,38	65	22
Total	23,213	611	**103**

1) INCLUSION CRITERIA

The inclusion criteria is based on the below terms:

- Those research papers will be extracted where big data is discussed in detail.
- Those papers will be included where a detail discussion is present about cloud computing.
- Those papers which discuss about cloud computing security challenges.
- Those research papers which discussed about big data security challenges on the platform of cloud computing.
- Those papers which have discussed about the solutions/practices for big data security challenges on cloud computing.
- The papers that were written in the English language.
- The papers which have similar title to our research article.
- The paper which have the keywords similar to our defined search term.

2) EXCLUSION CRITERIA

The exclusion criteria is based on the following terms:

- Those papers will be excluded, which are not related to big data.
- Those papers will also be excluded, which are not relevant to cloud computing.
- Those papers will be excluded, which does not match our research questions.
- Those papers that have different title from our search string.
- Those papers, which do not match the abstract with our search term.
- Those paper that did not match the keywords with our research string.
- Duplicate papers will also be excluded.
- Those papers that are not written in English language.

3) PUBLICATION SELECTION & QUALITY ASSESSMENT

Publication selection is basically the criteria of selection of a research publication, which is done on the basis of paper title, abstract, and keywords. We have selected 611/23,213 papers primarily. The quality of the publications will be assessed when the publication will be selected finally. The quality assessment of a publication will depends on the below questions.

- Does the author clearly identify the issues faced by vendor organization that can affect security of big data usage on cloud computing?
- What are the practices that are adopted by the author to solve these security issues?

4) DATA EXTRACTION

For data extraction we have followed the below criteria's:

- Details of paper publication i.e. Title, Authors Name, Information about Reference Type that either the paper is published in Journal or Conference, Conference Name, Journal Name, Issue and Volume of the Journal, Location of Conference, Year of Publication, Number of Pages etc.
- That data which is associated with our research questions.

5) DATA SYNTHESIS

With the help of SLR, we identified a list of the security challenges from the sample of 103 published relevant research papers that are included in the SLR. Data synthesis has been performed mostly by the first author of the paper while a handsome support was provided by the rest of authors on the validation of the SLR results. Initially we identified 19 security challenges for the vendor organizations when dealing with cloud data platform, that are further reviewed and validated and therefore some of these security challenges were merged together on the basis of similarities such as geographical data location and distribution data storage were merged together as one challenge. Finally we have a list of 15 challenges which are shown in Table 2.

IV. FINDINGS

This section describes in detail the results obtained from the SLR.

A. CHALLENGES/BARRIERS/ISSUES FIND THROUGH SYSTEMATIC LITERATURE REVIEW

To answer RQ1, we have identified critical security challenges for vendor organizations by critically analysing research papers review through SLR that are shown in Table 2. The critical security challenges of big data usage on the platform of cloud computing are identified along with their occurrences in each research paper included in the SLR are: Data Secrecy Issue with (97%), means that the data secrecy issue has been discussed in 100 research papers included in the SLR, Geographical data location issue

with (69%), Unauthorized data access Issue with (65%), Lack of Control with (60%), Lack of Data Management having (59%), Network level issues having (58%), Data integrity issues with (56%), Data Recovery issues with (55%), Lack of Trust with (54%), Data Sharing Issue with (53%), Data Availability with (47%), Assets Issue with (35%), Legal Amenabilities with (33%), Lack of quality issues with (25%), and Lack of consistency with (25%). The cloud computing has still security concerns at each stage and on different viewpoints. According to an author from vendors perspective cloud computing has still no security as a service on its best. In cloud, users applications are organized and running on various virtual machines, there is possibilities of providing any vulnerabilities by the cloud vendors organization, which may exist in cloud services, operating systems, or user application can be affected by the hackers [48], [79].

The Geographical data location issue is reported in 69 percent research articles as a possible security challenge that might have a negative impact on the security of big data on cloud computing. In cloud computing, for large amount of data retrieval and storage lot of solutions have been proposed, although many of them have been applied in cloud computing but still there exits many issues that can delay the effective implementation of these solutions. These includes the capacity of cloud technologies and high performance for addressing a large volume of data, enhance the existing file systems for the demand of volume of data retrieval applications, and data storage that how can data will be easily extracted and transfer among the servers [49]. In cloud computing, customer data is stored at several locations all around the globe, where users don't have any idea that at what exact location their data is stored. In literature, there are many issues linked with the data storage, which are challenges related to ''shared storage media'', challenges related to ''location of data'', and challenges associated with ''reliability of storage media'' [15].

The unauthorized data access issue challenge is highlighted in 65 percent research articles of our research work. The data access issue occurs when a user accessing its own data or organizational data. In case of large scale organization of cloud computing, only the relevant employees are given the access to the confidential data, but according to the security policy of cloud and free access of user, an unauthorized user can also get access to that confidential data, which is a sever issue for the organization [16]. According to CERT Insider Threat Centers [50] malevolent insider are users that have an authorized access to the organizational data, services, or systems and deliberately misuses the data or services, which can affect organizational security, confidentiality, availability, or integrity.

The Lack of Control security challenge is described as 60 percent of research articles. The cloud organizations don't have a direct control on the data and also have no information about their data usage by someone else, which can cause a security issue since there is no translucent technique by which we can directly observe all the resources. It is also possible that data may not be completely removed from server at the time of data sharing [51], [62], [65], [75], [79].

The issue of ''Data management'' has been recorded in 59 percent of the relevant publications. The Big data is still facing the security issues when an organization moving it to cloud, as data management and analysis will be provided by the different providers [38].

The ''Network Level issue'' is reported in 58 percent research articles, which is a critical issue faced by software vendors organization while transferring the data over the network. The Network security deals with the security features and network protocol that are used for the transfer of data.

The ''Data integrity issues'' is defined in 56 percent of the published papers. The data integrity issue take place whenever a hidden modification occur in the data by some malicious insiders or accidental modification occurred so that receiver cannot see the exact data which the sender shared [52].

The ''Data Recovery issues'' is stated in 55 percent of the related research work. The data recovery is also a big issue for big data on cloud, as whenever there will be any data loss occur on cloud then recovery of complete data is not an easy task [11].

The ''Lack of Trust'' is also a critical issue for big data on cloud computing and reported in 54 percent of research articles. For cloud service providers, it is better to provide new data control policies to build trust between the user's [11], [74].

The ''Data Sharing Issue'' is mentioned and discussed in 53 research papers. The data sharing is a critical issue, which is faced by the big data with many sub-challenges i.e. data transferring speed and traffic jam. The transfer speed tells about the data transfer from one point to another while the traffic jam takes place between local sites, cities or world-wide during data sharing [53], [65].

The ''Data Availability'' is a critical issue which is reported in 47 percent research articles. The goal of data availability is the free access to the data and cloud resources [54].

The ''Assets Issue'' is a critical challenge for big data on cloud which is described in more than 35 percent of research papers. The co-occurrence of assets of multiple occupants at the same address, having lack of security control information's while using the same service of the cloud. The possibility of attacks increased while hosting the set of valued assets on obtainable infrastructure publicly [55].

The ''Legal Amenabilities is reported in 33 percent of published research papers. The guidelines and regulations, for example HIPA and SOX forbid the usage of cloud computing. In a variety of disciplines legal amenabilities are necessary for specific information technology infrastructures [56].

Our research findings can benefit the software vendors organization security about their big data usage on the platform of cloud computing. The list of identified security challenges from the literature are shown below in TABLE 2.

The TABLE 2 shows all the 15 critical security challenges along with their frequencies and percentages. All the iden-

TABLE 2. List of identified challenges through SLR.

S. No	Challenges	Frequency (N=103)	Percentage
1	Data Secrecy Issue	100	97%
2	Geographical data location issue	71	69%
3	Unauthorized data access Issue	67	65%
4	Lack of Control	62	60%
5	Lack of Data Management	61	59%
6	Network level issues	60	58%
7	Data integrity issues	58	56%
8	Data Recovery issues	57	55%
9	Lack of Trust	56	54%
10	Data Sharing Issue	55	53%
11	Data Availability	48	47%
12	Assets Issue	36	35%
13	Legal Amenabilities	34	33%
14	Lack of quality issues	26	25%
15	Lack of consistency	26	25%

tified critical security challenges have percentages >=25% which describes that all these security challenges have great impact on the big data security on cloud computing. Among these security challenges we have identified that the Data secrecy issue is the top most issue with the percentage of 97%, which big data faced on cloud platform. The software vendors companies can use our proposed research study to overcome these security issues on cloud platform for better protection of their data.

1) ANALYSIS OF THE SECURITY CHALLENGES FOR BIG DATA USAGE ON CLOUD COMPUTING, IDENTIFIED THROUGH SLR, FOR SOFTWARE VENDORS ORGANIZATION

We performed a statistical analysis on the identified challenges based on four different variables. These variables includes, different continents, time decade, research methods, and database used in the paper. The objective of these analyses is to identify that whether these security challenges remain the same as in every continent, time decade, research method, and database respectively or vice versa.

a: ANALYSIS AND COMPARISON OF CHALLENGES/BARRIERS ACROSS VARIOUS SEARCH ENGINES

We have compared several security challenges identified in two time decades (2000-2010, and 2011-2020). To answer RQII, Table 3 shows a list of challenges identified in these two decades. For the analysis of identified security challenges we have adopted linear by linear association Chi- square test, to find out if there is any significant differences between the challenges in these two decades. The Chi-square linear by linear association is considered more powerful than that of Pearson Chi-square test [57]. When the value of $p < 0.05$ we usually refer this difference as a significant difference. In case of Data recovery and Data sharing issues the value of $p < 0.05$, which indicates that in first decade

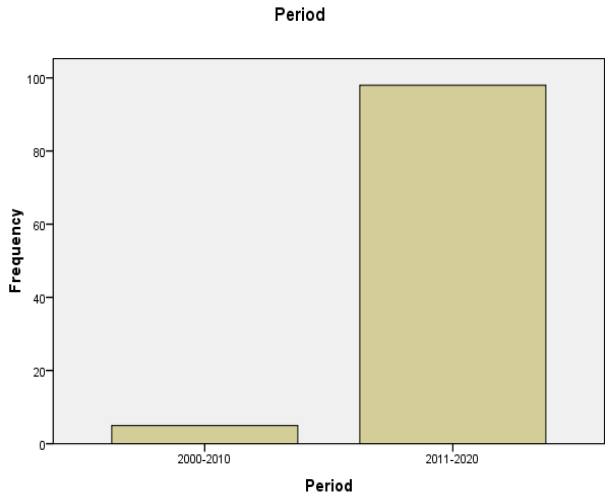

FIGURE 1. Final collection of papers from different time periods, identified through SLR.

(2000-2010) software vendors organization did not faced any such issue because of low trend of cloud for big data usage. It shows that these challenges are not critical in first decade. But if we see these two challenges in the other decade i.e. from (2011-2020), we can see that these two challenges have frequencies of 57 and 55 respectively, and a percentage of 58% and 56%. So in this two decades data recovery issue and data sharing issue have a significance difference. Furthermore, we have seen that data secrecy issue is the most critical issue for big data in both the decades as in decade (2000-2010) it has the percentage of 100% and in decade (2011-2020) the percentage is 97%. Similarly, the geographical data location issue has the percentage of 100% in decade (2000-2010) and 67% in decade (2011-2020), which shows that this security issue is also critical in both the decades. The lack of control issue also has 100% in first decade and 58% in the second decade, which means that this issue is also critical to overcome while using big data on cloud computing.

We have identified total of 15 challenges from the research articles for these two different decades. Our findings reveal that ''Data Secrecy issue'', ''Geographical data location issue'', ''Unauthorized data access Issue'', ''Lack of Control'', ''Lack of Data Management'', ''Network level issues'', ''Data integrity issues'', ''Data Recovery issues'', ''Lack of Trust'', ''Data Sharing Issue'', ''Data Availability'', ''Assets Issue'', and ''Legal Amenabilities'' are considered as critical challenges in these two different decades.

The core objective of this research study is to scratch different challenges which have negative influence on software vendor organizations for big data security over cloud computing. The identified security challenges across two different decades are shown in TABLE 3 (RQII).

In Figure 1, we have analyze the frequencies of the security challenges of big data on cloud computing platform in two different time period 2000-2010 and 2011-2020, which described that in in first decade only in 5 research articles these security challenges were highlighted while in second decade there are 98 published articles where these security challenges were discussed. Which clearly shows that these security challenges are very much critical in the second (decade 2011-2020).

b: ANALYSIS AND COMPARISON OF BIG DATA SECURITY ON CLOUD COMPUTING, IDENTIFIED THROUGH SLR, BASED ON DIFFERENT CONTINENTS

We have used SPSS tool in order to find out the frequencies of different security challenges of big data on cloud computing across different continents. To analyze these security challenges we have used chi-square linear by linear association test to find out the significance difference among these challenges across various continents. We compared these challenges across six different continents (Asia, Europe, North America, South America, Africa, and Australia), and also a mixed (combination of two or more continents i.e. Africa and Australia). We have find out the similarities and dissimilarities of these challenges across different continents. The Details of Risk/Challenges across different Continents are shown in TABLE 4.

We have found only one significant difference for the challenge ''Unauthorized data access issue'' across these continents as mention in the below Table 4. Where this challenge has not found in South America (0%) and (42%) in that of North America, jointly have (27%). This means that unauthorized data access issue is not critical for South America. In Africa reported as 50%, Europe 57%, mixed 67%, and in Asia it was reported in 77% articles. Which shows that for all other continent except South America the issue of un-authorized access was consider as critical challenge.

Data availability issue has (0%) frequency in Australia which means that this security is not critical in this continent. While in Africa having a highest occurrences of 100% in research articles means for Africa this security challenge is very critical.

The percentage of ''Legal Amenabilities'', ''Lack of quality issues'', and ''Lack of consistency'' were reported (0%) in Australia which shows that these security issues were not critical in this continent. Moreover our findings shows that these challenges are critical for the rest of the continents and mixed one.

In Figure. 2 we have described about the frequency analysis of our identified security challenges across different continents, these are consists of Asia, Europe, North America, South America, Africa, Australia, and Mixed continent of Africa and Australia, where we have identified that Asia is on top among these continents which shows that in Asia these challenges are very much critical for big data usage on cloud computing.

c: ANALYSIS OF BIG DATA SECURITY ON CLOUD COMPUTING, IDENTIFIED THROUGH SLR, BASED ON DIFFERENT STUDY STRATEGY (METHODS)

TABLE 5, describes the results identified through the SLR based on the study strategy. The sample size for this study

TABLE 3. List of identified security challenges through SLR in two different decades.

Challenges	Sample size find through SLR (n=103) Period				Chi- square test (linear by linear association) a=.05		
	2000-2010 (n=5)		2011-2020 (n=98)		X^2	Df	P
	Freq	%	Freq	%			
Data Secrecy Issue	5	100	95	97	.156	1	.693
Geographical data location issue	5	100	66	67	2.346	1	.126
Unauthorized data access Issue	2	40	65	66	1.436	1	.231
Lack of Control	5	100	57	58	3.441	1	.064
Lack of Data Management	2	40	58	60	.713	1	.398
Network level issues	2	40	58	60	.713	1	.398
Data integrity issues	3	60	55	56	.029	1	.865
Data Recovery issues	0	0	57	58	6.45	1	.011
Lack of Trust	4	80	52	53	1.378	1	.240
Data Sharing Issue	0	0	55	56	5.963	1	.015
Data Availability	3	60	45	46	.375	1	.540
Assets Issue	1	20	36	37	.573	1	.449
Legal amenabilities	1	1	34	35	2.564	1	.109
Lack of quality control process	1	20	25	26	.076	1	.783
Lack of consistency	2	40	24	24	.601	1	.438

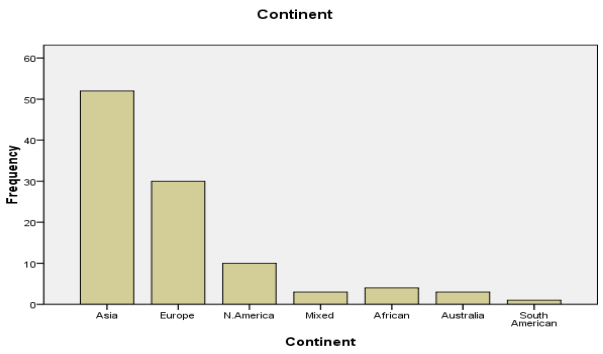

FIGURE 2. Final selection of papers from different continents, identified through SLR.

contains 103 research papers that are identified through the SLR.. We have used the SLR protocol to extract data from each research paper. For our sample size, we have used six study strategies which includes Interviews, Surveys, SLR's, Literature reviews, Experience reports, Thesis reports. Below is the Table 5, which describes the identified challenges through SLR across various methods/strategies. Figure 3 describes the different articles with respect to different methods used in this research paper.

From our findings, we concluded that majority of the challenges were highlighted in the literature reviews, surveys, and experience reports. In case of interviews, there are seven most critical challenges having the frequencies 100% in research articles they are 'Data secrecy issue', 'Geographical data access issue', 'Lack of control', 'Lack of data management', 'Data sharing issue', 'Data availability issue', and 'Lack of data consistency'. While the rest eight challenges are not properly discussed or might be out of scope in case of interview as they are highlighted in 0% published papers. These are 'Unauthorized data access Issue', 'Data integrity issue', 'Data recovery issue', 'Lack of Trust', 'Assets issue', 'Legal amenabilities', and 'Lack of quality issue'.

In case of Survey's, all the challenges were critical except 'Lack of quality control issue' which is reported in 17% research articles. 'Data secrecy issues' is the most repeated challenges in this strategy with a percentage of 97%. The rest of the challenges have percentage greater than 25%. In SLR, all the challenges are describes in 100% published articles except 'Lack of quality control', and 'Lack of data consistency' which have the percentage of 0% in the related articles.

In the strategy of Literature review 'data secrecy issue' is the most reported challenge having a percentage of 97%. While 'Lack of data consistency' is described in 19% related article which shows that this issue was not critical for the method of Literature review.

TABLE 4. List of challenges identified through SLR across different continent.

CHALLENGES	Sample size find through SLR (n=103) continent												X^2	df	p
	Asia (n=52)		Europe (n=30)		N&S America (n=11)		Australia (n=3)		Africa (n=4)		Mixed (n=3)				
	F	**%**	**F**	**%**	**F**	**%**	**F**	**%**	**F**	**%**	**F**	**%**			
Data Secrecy Issue	51	98	28	93	11	100	3	100	4	100	3	100	.009	1	.9
Geographical data location issue	38	73	19	63	8	73	1	34	4	100	1	34	.43	1	.51
Unauthorized data access Issue	**40**	**77**	**17**	**57**	**3**	**27**	**3**	**100**	**2**	**50**	**2**	**67**	**3.9**	**1**	**.05**
Lack of Control	34	65	15	50	5	45	3	100	2	50	3	100	.004	1	.95
Lack of Data Management	32	62	20	67	4	37	1	34	2	50	1	34	1.2	1	.26
Network level issues	31	60	19	64	3	27	2	67	3	75	2	67	.18	1	.67
Data integrity issues	31	60	13	44	6	55	3	100	4	100	1	34	.137	1	.71
Data Recovery issues	33	64	16	54	3	27	3	100	2	50	0	0	2.002	1	.16
Lack of Trust	30	58	14	47	6	55	1	34	3	75	2	67	.317	1	.57
Data Sharing Issue	27	52	18	60	6	55	2	67	2	50	0	0	.030	1	.86
Data Availability	26	50	13	44	3	27	0	0	4	100	2	67	.030	1	.86
Assets Issue	15	29	13	44	4	37	1	34	3	75	1	34	1.465	1	.23
Legal amenabilities	17	33	12	40	2	18	0	0	2	50	1	34	.300	1	.58
Lack of quality control process	12	23	8	27	3	27	0	0	1		2	67	.014	1	.91
Lack of consistency	10	19	9	30	3	27	0	0	2	50	2	67	2.432	1	.12

*F= Frequency

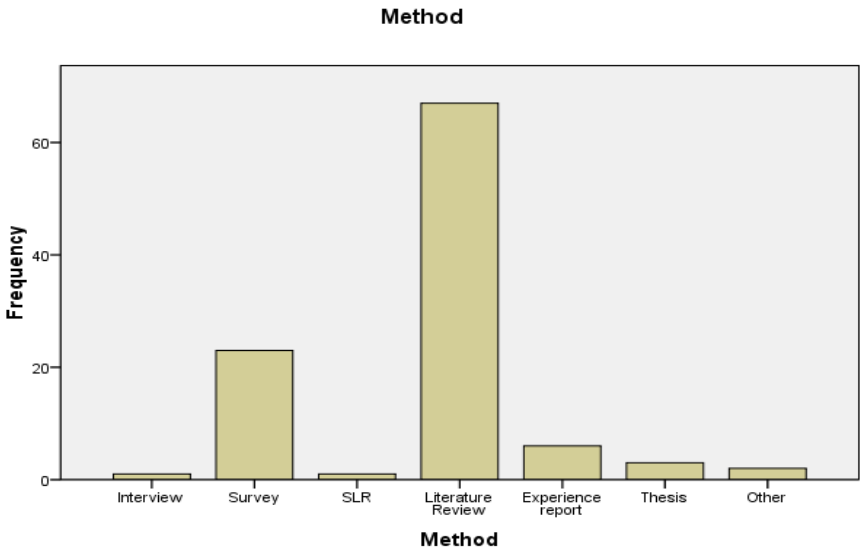

FIGURE 3. Final collection of papers from different strategies, identified through SLR.

The rest 15 challenges are critical because all have their percentages great than 25% as shown above in TABLE 5. 'Data secrecy issue' is the most critical challenges for both experience report and thesis repot as it is reported 100% in published research work for both the strategies. The issue 'Lack of quality control process' described in 0% research articles, and the challenges 'Network level issue', 'Assets issue', 'Legal amenabilities', 'Lack of data consistency' were recorded in 17% published work in case of strategy of experience report. In the strategy of thesis report the issue of 'Data

TABLE 5. List of challenges identified through SLR across different methods.

Challenges	Sample size find through SLR (n=103) Method														X^2	df	p
	Inter .v (n=1)		Survey (n=23)		SLR (n=11)		L. review (n=3)		Report (n=6)		Thesis (n=3)		Others (n=2)				
	F	%	F	%	F	%	F	%	F	%	F	%	F	%			
Data Secrecy Issue	1	100	22	96	1	100	65	97	6	100	3	100	2	100	.3	1	.6
Geographical data location issue	**1**	**100**	**18**	**78**	**1**	**100**	**45**	**67**	**4**	**67**	**2**	**67**	**0**	**0**	**4.6**	**1**	**.03**
Unauthorized data access Issue	0	0	15	65	1	100	45	67	3	50	2	67	1	50	.02	1	.9
Lack of Control	1	100	15	65	1	100	41	61	2	33	1	33	1	50	1.7	1	.2
Lack of Data Management	1	100	15	65	1	100	39	58	2	33	2	67	0	0	3.4	1	.06
Network level issues	0	0	16	70	1	100	39	58	1	17	2	67	1	50	1.1	1	.289
Data integrity issues	0	0	13	56	1	100	38	57	3	50	3	100	0	0	.12	1	.725
Data Recovery issues	0	0	10	43	1	100	41	61	3	50	1	33	1	50	.5	1	.480
Lack of Trust	0	0	13	56	1	100	37	55	3	50	1	33	1	50	.08	1	.781
Data Sharing Issue	1	100	13	56	1	100	36	54	3	50	0	0	1	50	1.2	1	.270
Data Availability	1	100	13	56	1	100	28	42	2	33	3	100	0	0	1.9	1	.161
Assets Issue	0	0	9	39	1	100	22	33	1	17	3	100	1	50	.25	1	.619
Legal amenabilities	0	0	7	30	1	100	22	33	1	17	3	100	0	0	.04	1	.837
Lack of quality control process	0	0	4	17	0	0	21	31	0	0	1	33	0	0	.04	1	.848
Lack of consistency	**1**	**100**	**10**	**43**	**0**	**0**	**13**	**19**	**1**	**17**	**1**	**33**	**0**	**0**	**5.4**	**1**	**.020**

TABLE 6. List of challenges identified through SLR across different databases.

Challenges	Sample size find through SLR (n=103) Database										X^2	df	p
	IEEE (n=11)		ACM(n=4)		Science Direct (n=1)		Springer Link (n=4)		Google Scholar (n=3)				
	F	%	F	%	F	%	F	%	F	%			
Data Secrecy	10	91	4	100	1	100	20	91	65	100	2.2	1	.14
Geographical data location	5	45	3	75	0	0	14	64	49	75	3.8	1	.05
Unauthorized data access	**9**	**82**	**4**	**100**	**1**	**100**	**15**	**68**	**38**	**58**	**4.3**	**1**	**.04**
Lack of Control	5	45	3	75	1	100	11	50	42	65	.68	1	.41
Lack of Data Management	4	36	2	50	0	0	17	78	37	57	1.7	1	.18
Network level issues	6	55	3	75	0	0	15	68	36	55	.02	1	.89
Data integrity issues	6	55	2	50	1	100	10	45	39	60	.13	1	.71
Data Recovery issues	7	64	0	0	1	100	8	36	41	63	.72	1	.39
Lack of Trust	8	73	1	25	0	0	8	36	39	60	.001	1	.98
Data Sharing Issue	5	45	1	25	0	0	11	50	38	58	1.8	1	.18
Data Availability	3	27	2	50	0	0	9	41	34	52	2.3	1	.12
Assets Issue	**1**	**9**	**1**	**25**	**0**	**0**	**9**	**41**	**26**	**40**	**4.2**	**1**	**.04**
Legal amenabilities	**1**	**9**	**0**	**0**	**0**	**0**	**11**	**50**	**22**	**34**	**3.9**	**1**	**.04**
Lack of quality control process	3	27	2	50	0	0	4	18	17	26	.14	1	.70
Lack of consistency	1	9	2	50	1	100	5	23	17	26	.19	1	.65

*F = Frequency

TABLE 7. Practices to address data secrecy issue.

CSC-1: Data Secrecy Issue Data secrecy issue is all about data security, privacy and confidentiality. The Protection of data is regularly a sequence of reactions and not a stratagem, Not being sure what is practical protection for different types of data, Not protecting important data appropriate to its value, Not knowing who uses what sensitive data					
S. No	**Practices**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC-1.1	Security and Privacy can be resolved using data encryption. The main technology for big data security is the technology of data encryption.	P-1,P-3,P-4,P-5,P-7,P-8,P-9,P-11,P-13,P-14,P-15, P-16, P-18,P-19,P-21,P-23,P-24, P-26, P-27, P-29, P-36, P-39, P-40, P-41, P-43, P-46, P-3,P-54, P-58, P-59, P-61, P-64, P-67, P-70, P-72, P-73, P-76, P-78, P-80, P-81, P-88, P-89, P-93, P-96, P-97, P-99, P-101, P-103	Survey= 19 SLR= 1 L. Review= 24 Report= 1 Thesis= 2 Interview= 1	48	47
PCSC-1.2	For ensuring the privacy of sensitive data anonymization approach can be used.	p-11, p-3, p-53	L. Review= 1 Survey= 1 Other= 1	3	3
PCSC-1.3	For better security and privacy the data protection could be used as a service.	P-1,P-3,P-4,P-5,P-7,P-8,P-9,P-11,P-13,P-14,P-15, P-16, P-18,P-19,P-21,P-23,P-24, P-26, P-27, P-29, P-36, P-39, P-40, P-41, P-43, P-46, P-53,P-54, P-58, P-59, P-61, P-64, P-67, P-70, P-72, P-73, P-76, P-78, P-80, P-81, P-88, P-89, P-93, P-96, P-97, P-99, P-101, P-103	Report= 1 Thesis= 2 Interview= 1 Survey= 19 SLR= 1 L. Review= 24	48	47
PCSC-1.4	Apache Knox (Gateway) provides border security to that of confidential access to Hadoop bunches along with organizational policies within companies	P-1,P-3,P-4,P-5,P-7,P-8,P-9,P-11,P-13,P-14,P-15, P-16, P-18,P-19,P-21,P-23,P-24, P-26, P-27, P-29, P-36, P-39, P-40, P-41, P-43, P-46, P-53,P-54, P-58, P-59, P-61, P-64, P-67, P-70, P-72, P-73, P-76, P-78, P-80, P-81, P-88, P-89, P-93, P-96, P-97, P-99, P-101, P-103	L. Review= 24 Interview= 1 Survey= 19 SLR= 1 Report= 1 Thesis= 2	48	47
PCSC-1.5	Data classification is also key for sensitive data identification and protection.	P-5,P-7, P-70	Survey= 1 L. Review= 2	3	3
PCSC-1.6	For providing information to cloud users related to the status of security of a cloud vendors the tool of security metrics is used.	P-1,P-3,P-4,P-5,P-7,P-8,P-9,P-11,P-13,P-14,P-15, P-16, P-18,P-19,P-21,P-23,P-24, P-26, P-27, P-29, P-36, P-39, P-40, P-41, P-43, P-46, P-53,P-54, P-58, P-59, P-61, P-64, P-67, P-70, P-72, P-73, P-76, P-78, P-80, P-81, P-88, P-89, P-93, P-96, P-97, P-99, P-101, P-103	Survey= 19 SLR= 1 L. Review= 24 Report= 1 Thesis= 2 Interview= 1	48	47
PCSC-1.7	The tools of Cryptography are also very much necessary for big data security protection techniques.	P-7, P-99, P-101, P-103,	Survey= 1 L. Review= 2 Thesis= 1	4	4

sharing' is highlighted in 0% articles. For the main difference among several study strategies, we have used linear by linear Chi-Square test in order to find out the key difference among the 15 big data critical security challenges on cloud.

Our findings describes that in all these strategies 'Data secrecy issue' is the mostly reported challenge that can affect software vendors organization while placing big data on cloud platform. Our result discloses more resemblances than that of its differences.

Our findings also tells about the importance of different strategies used in this literature. It also describes that which method is more suitable for our research work.

d: ANALYSIS OF BIG DATA SECURITY ON CLOUD COMPUTING, IDENTIFIED THROUGH SLR, BASED ON DIFFERENT DATA BASES

TABLE 6, describes about the different databases which we used in our research for relevant results and data extraction. We have used five different databases in our research process, which are: ''IEEE Explore, ACM, Science Direct, Springer Link,'' and Google Scholar. In which, we see that majority of the problems are accumulated from Google Scholar. The security issues 'Assets issues', 'Legal amenabilities', and 'Lack of data consistency' are reported in 9 percent of related articles in IEEE Explore, which shows that these are not critical in case of this database.

'Data secrecy' and 'un-authorized data access' issues are the most critical challenges in IEEE Explore as 91% and 82% respectively. 'Data recovery' and 'legal amenabilities' are reported in 0% research articles in ACM database, while Data Secrecy issue and un-authorized data access issues are the most critical challenges for the ACM database and both are reported 100% in the relevant published work.

e: PRACTICES TO ENHANCE DATA SECRECY ISSUE

In TABLE 7, we have addressed the practices to enhance data secrecy issue.

The challenges 'Data secrecy issue', 'un-authorized data access issue', 'Lack of control', 'Data integrity issue', 'Data recovery issue', and 'Lack of data consistency' are the most critical challenges reported in 100% published research work in database of Science Direct. The rest of all challenges are not critical for this database as these are reported 0% in research articles. For the database of Springer Link 'Lack of quality control issue', and 'Lack of data consistency' are the described in 18% and 23% respectively, which shows that these are not critical challenges. The remaining 13 challenges are critical for this database and reported in more than 35%

TABLE 8. Practices for enhancing geographical data location issue.

CSC-2: Geographical data location issue In cloud computing customer have no information about their data location, which create many issues. geographical distance, time zone difference etc.					
S. No	**Practices**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC-2.1	As in data classification data is sorted in various categories like type, and form etc. which can help to find out that which data is the most sensitive and where it is the storage location of that sensitive data.	P-4,p-8, p-11, p-14, p-5, p-16, p-18, p-21, p-23, p-24, p-26, p-27, p-31, p-35, p-36, p-53, p-61, p-76, p-78, p-79,	Survey= 9 L. Review= 8 SLR= 1 Report= 2	20	19
PCSC-2.2	Homomorphic encryption: With the usage of this technique user can stored the secret data on the cloud platform in the form of cipher text form and can be execute this secret data for any type of a necessary computations without the decryption of the cipher-text.	P-4,p-8, p-11, p-14, p-5, p-16, p-18, p-21, p-23, p-24, p-26, p-27, p-31, p-35, p-36, p-53, p-61, p-76, p-78, p-79,	L. Review= 8 Repot= 2 Survey= 9 SLR= 1	20	19
PCSC-2.3	Vulnerability: The vulnerability of the data may be reduced by evading the use of the open source software's in the virtualization servers.	p-4, p-15, p-27, p-79	Survey= 2 L. Review= 2	4	4
PCSC-2.4	In HDFS initially the stored data is encrypted from User side in the form of small chunks by using the parallel computation.	P-4,p-8, p-11, p-14, p-5, p-16, p-18, p-21, p-23, p-24, p-26, p-27, p-31, p-35, p-36, p-53, p-61, p-76, p-78, p-79,	Repot= 2 Survey= 6 SLR= 1 L. Review= 11	20	19
PCSC-2.5	The Existing cloud stowage solutions may deliver a decent level of security so that to protect customers secret data,	P-4,p-8, p-11, p-14, p-5, p-16, p-18, p-21, p-23, p-24, p-26, p-27, p-31, p-35, p-36, p-53, p-61, p-76, p-78, p-79,	L. Review= 7 Repot= 2 Survey= 10 SLR= 1	20	19
PCSC-2.6	In Cloud computing the various data copies may be saved at multiple places in order to maintain the security of the data	P-4,p-8, p-11, p-14, p-5, p-16, p-18, p-21, p-23, p-24, p-26, p-27, p-31, p-35, p-36, p-53, p-61, p-76, p-78, p-79	Survey= 9 L. Review= 8 SLR= 1 Report= 2	20	19

research papers. The issue of data secrecy is the most critical issue as highlighted in 91% research papers. in the database of Google scholar all the challenges are reported in more than 30% published work except 'Lack of quality control' and 'Lack of data consistency issue' both have 26% occurrences in research articles. 'Data secrecy issue' is the most critical challenge as reported in 100% of research work. For the main difference among various databases, we have used linear by linear Chi-Square test in order to find out the significant difference among the 15 big data security challenges on cloud computing. Our result discloses more comparations than its differences. Our findings also express about the importance of different databases used in this literature. It also tells about which database is more suitable in searching of our research paper. We have found the key differences for the issues of 'Un-authorized data access issue', 'Legal Amenabilities', and 'Assets issues' across different databases. Below is the table for identified challenges through SLR across various databases [7], [11], [24], [33], [40], [41].

2) PRACTICES IDENTIFIED THROUGH SLR, FOR ADDRESSING THE IDENTIFIED BIG DATA SECURITY CHALLENGES ON CLOUD COMPUTING

With the help of SLR we have identified a total of 63 practices for the security of big data usage on cloud computing from vendors perspective. Which are discussed in detail below:

In the below Tables of solutions/practices we have used several abbreviations. Which are:

- 'CSC' used for ''Critical Security Challenge''
- 'PCSC' used for ''Practices for Critical Security Challenge''
- 'P' used for paper, like P-1 denotes paper-1.

a: PRACTICES TO IMPROVE GEOGRAPHICAL DATA LOCATION ISSUE

See Table 8.

b: PRACTICES TO AUGMENT UN-AUTHORIZED DATA ACCESS ISSUE

See Table 9.

c: PRACTICES TO ENHANCE LACK OF CONTROL ISSUE

See Table 10.

d: PRACTICES TO AUGMENT LACK OF DATA MANAGEMENT ISSUE

See Table 11.

e: PRACTICES FOR IMPROVING NETWORK LEVEL ISSUE

See Table 12.

f: PRACTICES FOR AUGMENTING THE DATA INTEGRITY ISSUE

See Table 13.

TABLE 9. Practices to augment un-authorized data access issue.

CSC-3: Un-authorized data access issue
The issue in which an Un-authorized user can got access to someone personal data from an illegal approach or any other technique then this illegal access is called un-authorized data access issue.

Practice No	Practice	Paper-Id	Methods	Freq	%
PCSC3.1	For the identification of un-authorized customers, usage of the credential or attributed based policies are very much useful.	p-1, p-3, p-7, p-13, p-18, p-21, p-24, p-36,p-53,p-63, p-67, p-89, p-96, p-97, p-100	L. Review= 6 Survey= 5 Other=1 Reports= 3	15	15
PCSC3.2	For malware detection in cloud computing the technique of data mining is used.	p-3, p-36, p-97	L. Review= 2 Survey= 1	3	3
PCSC3.3	Data separation approach is used for illegal access prevention to the secret data in the cloud computing.	p-7, p-13, p-96	Survey= 1 L. Review=1 Thesis= 1	3	3
PCSC3.4	For the protection of the secret data from hackers the pro-active secret sharing technique can be used	p-1	L. Review= 1	1	1
PCSC3.5	A service named "Permissions as a service (PaaS)" can be used to inform the end-users that data may be accessed from which part or side.	p-97	L. Review= 1	1	1

TABLE 10. Practices to develop lack of control issue.

CSC-4: Lack of control issue
The lack of control issue produced by clouds makes it tougher while checking for the data integrity issue and confidentiality issue in such a situation. Uploading any personal data on the cloud simply means that some loss of data control may be produced.

Practice No	Practice	Paper-Id	Methods	Freq	%
PCSC4.1	While enhancing the security of data, it is very much significant to deliver access control, authentication of the data, and authorization to the data which are stowed in the cloud platform.	p-4, p-8, p-13, p-14, p-16, p-18, p-23, p-24, p-31, p-36, p-43, p-46, p-53, p-58, p-61, p-70, p-72, p-73, p-76, p-88, p-94, p-101, p-103	Survey= 7 L. Review= 10 Other= 1 Report= 2 Thesis= 2 SLR= 1	23	22
PCSC4.2	With the help of real time access control the data on the cloud will be measured securely.	p-14	L. Review= 1	1	1
PCSC4.3	The Apache ACCUMULO could enable very granulated access control to important-value of pairs.	p-18, p-101	Survey= 1 L. Review= 1	2	2
PCSC4.4	Logging at each layer is the best approach to control malware access.	p-14, p-18, p-36, p-88, p-94	L. Review= 4 Survey= 1	5	5
PCSC4.5	For the protection of the data, adequate access control approach is very much necessary	p-4, p-24, p-61, p-7	Survey= 2 L. Review= 1 Reports= 1	4	4
PCSC4.6	The data could be conserved with the help of cryptography technique and access control mechanism at every layer.	p-103	L. Review= 1	1	1

g: PRACTICES TO DEVELOP DATA RECOVERY ISSUE
See Table 14.

h: PRACTICES TO REPORT THE LACK OF TRUST ISSUE
See Table 15.

i: PRACTICES TO IMPROVE DATA SHARING ISSUE
See Table 16.

j: PRACTICES TO IMPROVE DATA AVAILABILITY ISSUE
See Table 17.

k: PRACTICES TO ADDRESS ASSETS ISSUE
See Table 18.

l: PRACTICES TO ADDRESS LEGAL AMENABILITIES
See Table 19.

m: PRACTICES FOR ADDRESSING LACK OF QUALITY ISSUE
See Table 20.

n: PRACTICES TO IMPROVE LACK OF CONSISTENCY
See Table 21.

TABLE 11. Practice to improve lack of data management issue.

CSC-5: Lack of Data Management issue Management issue, uncertain planning & decisions problems, Absolute volume of data. Taking a sensitive method to management of the data. Lack of processes and systems, and also the Scrappy data possession.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC5.1	Identity and Access Management (IAM): IAM contains the construction, data management, and deletion of a digital distinctiveness.	p-4, p-9, p-36, p-39, p-40, p-53, p-59, p-78, p-81, p-101	L. Review= 6 Reports= 1 Survey= 3	10	8
PCSC5.2	The Cloud computing security has the main objectives of key management of data confidentiality.	p-40, p-81	L. Review= 2	2	2
PCSC5.3	Intrusion Management: An increasing provision class is the use of devoted cloud providers to evaluate relatively huge data sets for the identification of the sign of interruption.	p-40, p-53, p-59, p-78, p-81, p-101	L. Review= 3 Survey= 2 Thesis= 1	6	6
PCSC5.4	Management Solutions and Responsibility- The Management Solutions and imposing liability controls could support in sufficient security procedures of critical data management.	p- 4, p-81, p-101	L. Review= 2 Survey= 1	3	3

TABLE 12. Practices for improving network level issue.

CSC-6: Network level issues These are Identical IP Addresses, IP Address Enervation, DNS issues, Single Workplace Incapable to Link the related Networks, Incapable to Link to a Local Printer or file Shares, Limited Networks is Incapable for the Connection to the internet, which can Slow the Internet efficacy.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC6.1	In Cloud for the detection of threats, the prevention system of network based intrusion is used.	p-4, p-7, p- 18, p-19, p-20, p-21, p-23, p-24, p-27, p-29, p-31, p-34, p-39, p-43, p-46, p-54, p-59, p-61, p-78, p-88, p-97, p-103	L. Review= 9 Survey= 8 SLR= 1 Interview= 1 Reports= 2 Thesis= 1	22	21
PCSC6.2	The protocols of the Network layer security can be certifies the secure network communications at every layer where packages will route over the networks.	p-4, p-7, p-18, p-19, p-20, p-54, p-59, p-61	L. Review= 4 Survey= 3 Reports= 1	8	8
PCSC6.3	In cloud computing at the network and data levels the encryption can be done with the help of Crypto methods.	p-21, p-23, p-54, p-59, p-61, p-78, p-97	L. Review= 4 Survey=3	7	7
PCSC6.4	For providing solid authentication for server as well as client applications cloud uses a Kerberos secured network authentication where transfer of the password upon the network has no need.	p-39	L. Review= 1	1	1
PCSC6.5	Over the internet (broadband networks) a cloud computing vendors organization can provides all services to their end users.	p-4, p-7, p- 18, p-19, p-20, p-21, p-23, p-24, p-27, p-29, p-31, p-34, p-39, p-43, p-46, p-54, p-59, p-61, p-78, p-88, p-97, p-103	Interview= 1 Reports= 2 Thesis= 1 L. Review= 9 Survey= 8 SLR= 1	22	21

V. ANALYTICAL HIERARCHY PROCESS (AHP)

The process of choosing or identifying alternatives from the given set of factors on the basis of the preference from the decision-makers is known as multi-criteria decision-making approach. This procedure become complex when it includes multiple criteria, while MCDM method describes the findings based on conflicted criteria, e.g. benefit and cost criteria [72], [73].

For this MCDM problem, AHP method is very important, which is used by many researcher for prioritization and analysis [87]–[89]. To prioritize and analyze the identified security challenges of big data, we used the MCDM technique based on the AHP method. The AHP method is basically used for deciding relevant weight between multiple criteria, identification and prioritization of the challenges, and pair-wise comparison method for calculation of weights of the criteria within decision-making problems [83]–[85]. AHP has mainly three phases [86], [73], [87], which are:

1) Decomposition of the complex decision-making challenges into basic hierarchical structure.
2) Conclusion of priority-weights of challenges and their sub-ordinate challenges by pair-wise comparisons.

TABLE 13. Practices to enhance data integrity issue.

CSC-7: Data integrity issues In data integrity issue of cloud computing users can face checking constraints, design limitations, software bugs, constraints of foreign key, initialization and declaration of program, and compile time saneness checks.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC7.1	Data integrity checks may be achieved at user side or server side.	p-8, p-14, p-16, p-18, p-24, p-46, p-59, p-63, p-76, p-93	L. Review= 6 Survey= 4	10	10
PCSC7.2	For the secure computation and infrastructure of data integrity cloud vendor used the third party authentication approach.	p-14, p-16, p-18, p-24, p-46, p-59, p-63, p-76, p-93	L. Review= 5 Survey= 3 Thesis= 1	9	9
PCSC7.3	For the validation of data integrity Cong Wang proposed an mathematical approach for dynamically store data in the cloud.	p-76	L. Report= 1	1	1
PCSC7.4	To enforce the data integrity organizations needs to consider the usage of multiple applications or middle ware layers for imposing the data integrity.	p-18, p-24, p-46,	L. Report= 2 Survey= 1	3	3
PCSC7.5	To overcome data integrity issue and to support public data integrity verification cloud vendors can use the NEC Labs provable data integrity solutions (PDIS).	p-59	L. Report= 1	1	1

TABLE 14. Practices to improve data recovery issue.

CSC-8: Data Recovery issues The process in which the lost, unreachable, corrupted, formatted, damaged data are retrieved from a removable media or secondary storage files is called data recovery, when the data cannot be retrieved by ordinary normal way of retrieval.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC8.1	The data de-duplication is also a good solution to diminish backup and offline data storage capacities.	"p-11, p-13, p-18, p-21, p-29, p-39, p-43, p-59, p-96"	L. Report= 5 Survey= 4	9	9
PCSC8.2	For handling Data leakage issue the "watermarking technique can be used."	"p-11, p-13, p-18, p-21, p-29, p-39, p-43, p-59, p-96"	L. Review= 4 Survey= 4 SLR= 1	9	9
PCSC8.3	Data loss prevention (DLP) technique is used for ensuring the data, both in dynamic or static, observe to the strategies as demarcated by the vendor firms.	"p-11, p-13, p-18, p-21, p-29, p-39, p-43, p-59, p-96"	Survey= 4 SLR= 1 L. Review= 4	9	9
PCSC8.4	Backup and recovery: To overcome such type of issues, backup and data recovery are the key for ensuring of data and service offered by vendors organization.	"p-11, p-13, p-18, p-21, p-29, p-39, p-43, p-59"	L. Review= 4 Survey= 4	8	8
PCSC8.5	For the detection of data leakage the technique of Impeding data leakage (IDL) is used which is occurred in cloud by using the approach of swarm intelligence.	p-39, p-43, p-59, p-96	Survey= 2 L. Reports= 2	4	4

3) Verification of the consistency level of results.

We have the following equation of AHP to prioritized and analyze the security challenges of big data on cloud platform.

$$A = \begin{Bmatrix} 1 & a12 \ldots . a & 1n \\ a21 & 1 \ldots . a & 2n \\ \ldots . & \ldots . & \ldots . \\ an1 & a\ n2 & 1 \end{Bmatrix} \text{ where } aij = 1/aji, \quad aij > 0.1 \tag{1}$$

$$AW\lambda max(W) \tag{2}$$

where A is a table of pair wise comparison for security challenges, λ max is the highest eigen vector value and W is the relevant weight.

$$CI = (\lambda max - n)/(n - 1) \tag{3}$$

$$CR = CI/\ RI \tag{4}$$

where n represent orders of the challenges, CI consistency index and RI is the consistency value of random index which have different values based on total numbers of challenges.

TABLE 15. Practices to address lack of trust issue.

CSC-9: Lack of Trust issue The organization has concerned with the trustworthiness and data storage location due to which they were less willing to use cloud for their critical data and sensitive business applications.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC9.1	Building trust between vendor and client is the way to overcome these security challenges as it creates entity relationships speedily and securely.	p-13, p-23, p-60, p-73, p-80, p-88	L. Review= 4 Survey= 2	6	6
PCSC9.2	For the communication in any of the distributed system there will be must a trust worthy relationship between the parties for correspondence	p-13, p-23, p-60, p-73, p-80, p-88	Survey= 2 L. Review= 3 Reports= 1	6	6
PCSC9.3	By applying standards cloud secrecy approaches, the vendors organization are continuously working to build their customers trust and to ensure about their information leakage security.	p-13, p-23, p-60	L. Review= 3	3	3
PCSC9.4	*(TCCP): To overcome with trust issue the proposed practice is known as the (TCCP) trusted cloud computing platform.*	p-73, p-80, p-88	Survey= 2 L. Review= 1	3	3

TABLE 16. Practices to augment data sharing issue.

CSC-10: Data Sharing issue Sharing of data into a form appropriate for the analysis is an difficulty in the espousal of big data. As data could also retrieved by somebody else using the similar network for data sharing.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC10.1	Data Sharing along with authenticated users is a main application for data management on cloud.	p-15, p-61, p-99	Survey= 1 L. Review= 2	3	3
PCSC10.2	To this issue data will be encrypted and combined with forged data to make secure stream for sharing.	p-15, p-61, p-99	Survey= 1 L. Review= 2	3	3

TABLE 17. Practices to enhance data availability issue.

CSC # 11 Data Availability issues Data availability is the issue which deals with the timeliness and consistency of the access and usage of the critical data. Which consists of data approachability. For examples: files of any websites, which must persist reachable to avoid site interruption and disturbance of a package.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC11.1	The Availability Issue: which ensured the authorized clients have access to critical data whenever needed.	p-4, p-40, p-19	L. Review= 2 Reports= 1	3	3
PCSC11.2	The goal of data or service availability is to ensure the users who will use them when ever needed.	p-58, p-59, p-64, p-101	L. Review= 2 Reports= 1 Survey= 1	4	4
PCSC11.3	Due to the availability of cloud-based solutions the cost of the storage is intensely decreased.	p-4, p-40, p-19, p-58, p-59, p-64, p-101	L. Review= 5 Reports= 1 Survey= 1	7	7

The complete accepted values of consistency ratio CR is (0.10), but if the complete value of CR < 0.10 then the priority weights of the challenges are satisfactory and suitable. But if the whole value of CR > 0.10, then it is mandatory to repeat the evaluating procedure from phase-1 to recover consistency.

A. FINDINGS OF PAIR WISE COMPARISON, PRIORITY WEIGHTS AND CHECKING CONSISTENCY

In Table 23 and 24, we find out the pairwise comparison matrix by using the equation 1 and normalized matrix using equation 3, 4, and 5 for steadiness level. Where we have λ_{max} value of 3.029 used for the largest eigen value and CR value is.025, which is less than 0.1 which is satisfactory and acceptable.

Table. 25 and 26 describe about the pairwise comparison matrix and normalized matrix for the management level. Where the λ_{max} value is 5.313 and CR value is 0.07 less than 0.1. Thus, the priority weights of the challenges are satisfactory and acceptable. In case if the CR value is not less than 0.1, then reiterate the evaluating procedure for improving the consistency.

TABLE 18. Practices to improve assets issue.

CSC-12. Assets issues
Cloud computing become more advanced day by day for the organizational data storage over the internet as every organization are placing their data on cloud platform. Due to which every organization has facing tough time with the cloud tools and to up to date their experts experience about cloud latest technologies.

Practice No	Practice	Paper-Id	Methods	Freq	%
PCSC12.1	Management Resource techniques are very significant for PaaS and SaaS providers to support in the management of the type and quantity of possessions allotted to dispersed applications.	"p-1, p-19, p-20, p-27, p-34, p-40, p-53, p-78"	SLR= 1 L. Review= 5 Survey=2	8	8
PCSC12.2	The cloud based vendors can provides a pool of resources for example computing machines, storage devices and cloud networks to end users.	p-1, p-19, p-20, p-27, p-34, p-40	L. Review= 3 Survey= 2 Thesis= 1	6	6
PCSC12.3	Whenever we will implement cloud services and security on a global scale then it will be more cost effective	p-1, p-19, p-20, p-27, p-34, p-40, p-53, p-78	SLR= 1 L. Review= 5 Survey=2	8	8
PCSC12.4	Cloud computing offered environment to the development of agile and also minimize the cost, as cloud base BI can minimize the overall development costs.	p-1, p-19, p-20, p-27, p-34, p-40	L. Review= 3 Survey= 2 Thesis= 1	6	6

TABLE 19. Practice to improve legal amenabilities.

CSC-13: Legal Amenabilities
In today's issues of cloud computing is facing of the Amenability. Whenever an organization shift their data to cloud platform then the legal amenabilities were faced as yielding with organizations regulations and laws.

Practice No	Practice	Paper-Id	Methods	Freq	%
PCSC13.1	Many companies have misplaced the important amount of clients data due to data fissures. Organizations were trying to deliver regulations which can better avoid this issues.	p-18, p-36	L. Review= 2	2	2
PCSC13.2	For securing important data processing the sectoral legislation are used in cloud platform as an legislation and regulations.	p-18, p-36	L. Review= 2	2	2

TABLE 20. Practice to address lack of quality issue.

CSC-14: Lack of quality issues
Deprived data quality issue has become a solemn issue for several cloud service providers because data are often gathered from multiple sources.

Practice No	Practice	Paper-Id	Methods	Freq	%
PCSC14.1	Documentations supports vendor's quality distribution of goods. For better quality better documentation is essential	p-18	**L. Review= 1**	1	1
PCSC14.2	To overcome with quality issues service level agreement is used, which has the necessary to identify consequences reasons, and objectives of efficiency etc.	p-18	**L. Review= 1**	1	1
PCSC14.3	Follow standard testing procedures to enhance this issue.	p-18	**L. Review= 1**	1	1

Table 27 and 28 describe the pairwise comparison and normalized matrix by using the above mentioned equations for the control level of challenges. Here λ_{max} the largest eigen value of the Table 27 is 5.350 and CR value is 0.078 which is less than 0.1 so again the priority weights of the challenges are satisfactory and acceptable.

Table 29 and 30 figure out pairwise comparison and normalized matrix of eminence level with λ max 2.250 and CR value 0.063 less than 0.1 which shows that the priority weights are satisfactory and acceptable.

While, Table 31 and 32 describes the pairwise comparison and normalized matrix of these four levels, where λ_{max} has a value of 4.214, and CR is 0.079 which is less than 0.1. which described that the priority weights of these levels are satisfactory and acceptable.

In Table 33 we have discussed in detail about the local and global weights of the challenges and their ranks of priority.

TABLE 21. Practices to address lack of consistency issue.

CSC-15: Lack of consistency Cloud computing faces lack of data consistency issue for big data usage. This issue can be best avoided with consistent secrecy checks bridging on traditional servers and infrastructures of private cloud computing.					
Practice No	**Practice**	**Paper-Id**	**Methods**	**Freq**	**%**
PCSC15.1	In cloud computing the high quality of data can be characterized by data consistency.	p-54, p-70	L. Review= 1 Survey= 1	2	2
PCSC15.2	Cloud computing uses multiple websites which can improve the reliability and consistency, and make it possible for backup and business continuity.	p-54, p-70	L. Review= 1 Survey= 1	2	2

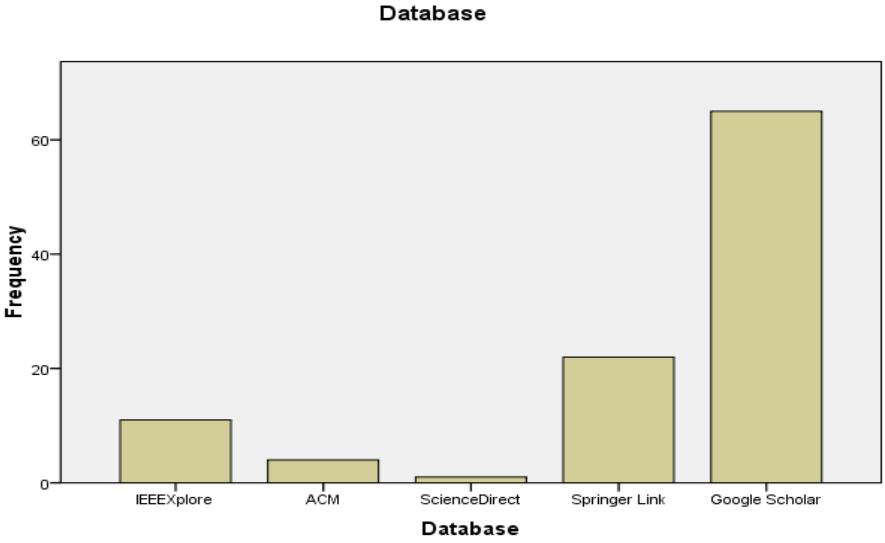

FIGURE 4. Final collection of papers from different databases, identified through SLR.

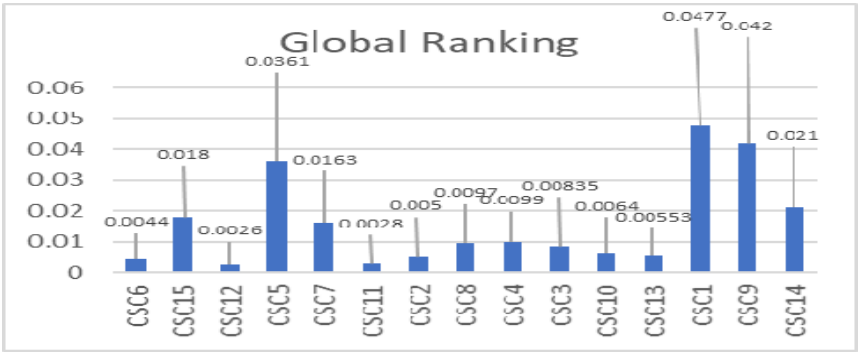

FIGURE 5. Local and global weight comparison of the challenges.

The global weights describes about the contribution of a specific challenge with in the comprehensive study. Global weights of the challenge is the product of local weights of the challenges and weights of its concerned level. For example the global weight for CSC is 0.048 which is obtained as 0.078*0.612. Similarly we have concluded the global ranking values for rest of the challenges described above. Figure 5 shows the overall global ranking of these challenges in detail. Among these 15 critical security challenges CSC1 has the highest global value of 0.0477, and CSC12 has the least global value of 0.00026. The rest of the challenges have the values greater than CSC12 and less than CSC1.

Table 34 described about the detail of prioritization and ranking of these critical security challenges of big data on the platform of cloud computing. Where CSC1 ''data secrecy issue'' has prioritized as the top most critical or significant

TABLE 22. Importance detail of SAATY basic 9 point scale.

AHP/SAATY Scale	Linguistic scale
1	Equally important (EI)
3	Weakly important (WI)
5	Fairly important (FI)
7	Strongly important (SI)
9	Absolutely more important (AMI)
2, 4, 6, 8	Intermediate values (IV)

TABLE 23. Pairwise comparisons for steadiness level challenges.

S#	Challenge	CSC6	CSC15	CSC12
1	CSC6	1.00	0.20	2.00
2	CSC15	5.00	1.00	6.00
3	CSC12	0.50	0.17	1.00
	Sum	6.50	1.37	9.00

TABLE 24. Normalized matrix for steadiness level challenges.

S#	Challenge	CSC6	CSC15	CSC12	Priority weights
1	CSC6	0.15	0.15	0.22	0.174
2	CSC15	0.77	0.73	0.67	0.723
3	CSC12	0.08	0.12	0.11	0.103
	Sum	1.00	1.00	1.00	1.000

λ max= 3.029, CI= 0.015, RI= 0.58, CR= 0.025<0.1

TABLE 25. Pairwise comparisons for management level challenges.

S#	Challenge	CSC5	CSC7	CSC11	CSC2	CSC8
1	CSC5	1.00	2.00	8.00	8.00	8.00
2	CSC7	0.50	1.00	5.00	4.00	2.00
3	CSC11	0.13	0.20	1.00	0.33	0.20
4	CSC2	0.13	0.25	3.00	1.00	0.33
5	CSC8	0.13	0.50	5.00	3.00	1.00
	Sum	1.88	3.95	22.00	16.33	11.53

TABLE 26. Normalized matrix for management level challenges.

S#	Challenge	CSC5	CSC7	CSC11	CSC2	CSC8	Priority weights
1	CSC5	1.00	2.00	8.00	8.00	8.00	0.517
2	CSC7	0.50	1.00	5.00	4.00	2.00	0.233
3	CSC11	0.13	0.20	1.00	0.33	0.20	0.040
4	CSC2	0.13	0.25	3.00	1.00	0.33	0.071
5	CSC8	0.13	0.50	5.00	3.00	1.00	0.138
	Sum	1.00	1.00	1.00	1.00	1.00	1.00

λmax= 5.313, CI= 0.078, RI= 1.12, CR= 0.07<0.1

TABLE 27. Pairwise comparisons for control level challenges.

S#	Challenge	CSC4	CSC3	CSC10	CSC13	CSC1
1	CSC4	1.00	2.00	1.50	2.00	0.10
2	CSC3	0.50	1.00	2.00	1.50	0.20
3	CSC10	0.67	0.50	1.00	2.00	0.10
4	CSC13	0.50	0.67	0.50	1.00	0.20
5	CSC1	10.00	5.00	10.00	5.00	1.00
	Sum	12.67	9.17	15.00	11.50	1.60

TABLE 28. Normalized matrix for control level challenges.

S#	Challenge	CSC4	CSC3	CSC10	CSC13	CSC1	Priority weights
1	CSC4	0.08	0.22	0.10	0.17	0.06	0.08
2	CSC3	0.04	0.11	0.13	0.13	0.13	0.04
3	CSC10	0.05	0.05	0.07	0.17	0.06	0.05
4	CSC13	0.04	0.07	0.03	0.09	0.13	0.04
5	CSC1	0.79	0.55	0.67	0.43	0.63	0.79
	Sum	1.00	1.00	1.00	1.00	1.00	1.00

λ max= 5.350, CI= 0.088, RI= 1.12, CR= 0.078<0.1

TABLE 29. Pairwise comparisons for eminance level challenges.

S#	Challenge	CSC9	CSC14
1	CSC9	1.00	0.20
2	CSC14	5.00	1.00
	Sum	6.50	1.37

TABLE 30. Normalized matrix for eminence level challenges.

S#	Challenge	CSC9	CSC14	Priority weights
1	CSC9	0.67	0.67	0.667
2	CSC14	0.33	0.33	0.333
	Sum	1.00	1.00	1.00

λ max= 2.250, CI= 0.250, RI= 4, CR= 0.063<0.1

TABLE 31. Pairwise comparisons between categories of challenges.

S#	Levels	Steadiness	Management	Control	Eminence
1	Steadiness	1	2	4	6
2	Management	0.5	1	2	4
3	Control	0.25	0.5	1	6
4	Eminence	0.2	0.25	0.2	1

TABLE 32. Normalized matrix for categories of challenges.

Levels	Steadiness	Management	Control	Eminence	Priority weights
Steadiness	1	2	4	6	0.492
Management	0.5	1	2	4	0.260
Control	0.25	0.5	1	6	0.189
Eminence	0.2	0.25	0.2	1	0.059

λ max= 4.214, CI=0.071, RI=0.9, CR=0.079<0.1 ∑=1.00

security challenge for big data on the cloud platform, so vendors must take serious action against this security challenge. If we compare the result of SLR with AHP approach then we have seen that according to SLR CSC1 is also most serious security challenge as out of 103 research articles in 97 articles CSC1 is identified as the most critical security challenge.

TABLE 33. Summary of local and global weights challenges and their rankings.

Category	Weights	Challenges	Local Weights	Local Ranks	Global Weights	Global Ranks
Steadiness	0.025	Network level issues CSC6	0.174	6	0.00435	13
		Lack of consistency CSC15	0.723	15	0.018075	5
		Asset issues CSC12	0.103	12	0.002575	15
Management	0.07	Lack of data management CSC5	0.517	5	0.03619	3
		Data Integrity issue CSC7	0.233	7	0.01631	6
		Data Availability CSC11	0.04	11	0.0028	14
		Geographical data location issue CSC2	0.071	2	0.00497	12
		Data Recovery issue CSC8	0.138	8	0.00966	8
Control	0.078	Lack of Control CSC4	0.127	4	0.009906	7
		Unauthorized data access issue CSC3	0.107	3	0.008346	9
		Data Sharing issue CSC10	0.082	10	0.006396	10
		Legal Amenabilities CSC13	0.071	13	0.005538	11
		Data Secrecy issue CSC1	0.612	1	0.047736	1
Eminence	0.063	Lack of Trust CSC9	0.667	9	0.042021	2
		Lack of quality issues CSC14	0.333	14	0.020979	4

*CSC= critical security challenge

TABLE 34. Prioritizing the challenges.

C#	Challenges	Priority
CSC1	Data Secrecy issue	1
CSC2	Geographical data location issue	12
CSC3	Unauthorized data access issue	9
CSC4	Lack of Control	7
CSC5	Lack of data management	3
CSC6	Network level issues	13
CSC7	Data Integrity issue	6
CSC8	Data Recovery issue	8
CSC9	Lack of Trust	2
CSC10	Data Sharing issue	10
CSC11	Data Availability	14
CSC12	Asset issues	15
CSC13	Legal Amenabilities	11
CSC14	Lack of quality issues	4
CSC15	Lack of consistency	5

CSC9 ''lack of trust issue'' is identified as the 2nd most significant or critical security challenge for big data usage on cloud computing, and CSC12 ''assets issue'' is reported as the least significant or critical issue among these security challenges. The rest are shown in the above Table 34. By similar method we can prioritized the identified practices against each identified challenges.

VI. LIMITATION

The authors of all these research studies were not supposed to report the sincere reasons that why these security issues have negatively influence on software vendors organization for big data usage on cloud. These may be that the majority of research studies were literature review, Surveys, and experience report which may be further subject to publication bias. With the increasing number of papers publication in big data security on cloud from vendor perspective, our SLR procedure may have dropped some related publications. Moreover, some search engines like Google Scholar could not gave access completely for paper extraction.

VII. CONCLUSION AND FUTURE WORK

Initially we have identified a list of 19 security challenges with the help of SLR in which we have merged some challenges and finally got 15 challenges which are shown in the Table 2. These 15 challenges ''Data Secrecy issue (97%)'', ''Geographical data location issue (69%)'', ''Unauthorized data access issue (65%)'', ''Lack of Control (60%)'' ''Lack of Data Management''(59%), ''Network level issues''(58%), ''Data integrity issues''(56%), ''Data Recovery issues''(55%), ''Lack of Trust''(54%), ''Data Sharing Issue''(53%), ''Data Availability''(47%), ''Assets Issue''(35%), ''Legal Amenabilities''(33%), ''Lack of quality control process (25%)'', and ''Lack of consistency (25%)'' are marked as critical security challenges for big data usage on cloud computing from software vendors' perspective. Also identified the standard 64 practices for these 15 critical security challenges from the selected literature. The software vendors organization needs to focus on these 64 practices to overcome these 15 critical security challenges while using big data on cloud computing.

The goal of our research is to give a protected way to software vendor's organization's for big data usage on cloud computing. The future work of our research is to validate the

identified security challenges and also to find out practices for these security challenges with the help of empirical study apart from the identified and discussed above. Furthermore, we plan to conduct a case study in the relevant software vendor's organization as in that of Capability Maturity Model Integration (CCMI) model [82], to identify each vendor organization level of our proposed security model and finally to assist them in using big data on cloud. Additionally, in future, we want to prioritized and analyzed these security challenges of big data on cloud platform by applying Fuzzy TOPSIS approach to identify the most important and critical challenges amongst the identified one.

REFERENCES

[1] S. Subbalakshmi and K. Madhavi, "Security challenges of big data storage in cloud environment: A Survey," *Int. J. Appl. Eng. Res.*, vol. 13, no. 17, pp. 13237–13244, 2018.
[2] A. Gholami and E. Laure, "Big data security and privacy issues in the CLOUD," *Int. J. Netw. Secur. Appl.*, vol. 8, no. 1, pp. 59–79, Jan. 2016.
[3] M. Chen, S. Mao, Y. Zhang, and V. C. M. Leung, *Big Data: Related Technologies, Challenges and Future Prospects*. Cham, Switzerland: Springer, 2014.
[4] C. S. Kruse, R. Goswamy, Y. Raval, and S. Marawi, "Challenges and opportunities of big data in health care: A systematic review," *JMIR Med. Informat.*, vol. 4, no. 4, p. e38, Nov. 2016.
[5] A. McAfee *et al.*, "Big data: The management revolution," *Harvard Bus. Rev.*, vol. 90, no. 10, pp. 60–68, 2012.
[6] S. V. V. Jambunathan, "A review on big data challenges and opportunities," *Int. J. Latest Technol. Eng., Manage. Appl. Sci.*, vol. 5, no. 9, p. 4, Nov. 2016.
[7] V. M. S. Singh and S. Srivastava, "The big data analytics with Hadoop: Review," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 4, no. 3, p. 5, 2016.
[8] U. Singh, N. K. Solanki, M. K. Varma, and T. Sevak, "A review on big data protection of Hadoop," in *Proc. 2nd Int. Conf. Commun. Electron. Syst. (ICCES)*, Oct. 2017, pp. 943–950.
[9] S. Harini, "A review of big data computing and cloud," *Int. J. Pure Appl. Math.*, vol. 118, no. 18, p. 8, 2018.
[10] A. A. Gangawane and A. Devi, "Big data security issues and challenges in cloud computing," *Asian J. Converg. Technol.*, vol. 1, no. 6, pp. 1–5, 2015.
[11] N. Sharma and M. Shamkuwar, "Big data analysis in cloud and machine learning," in *Big Data Processing Using Spark in Cloud*. Singapore: Springer, 2019, pp. 51–85.
[12] K. Kaur, A. Syed, A. Mohammad, and M. N. Halgamuge, "Review: An evaluation of major threats in cloud computing associated with big data," in *Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA)*, Mar. 2017, pp. 368–372.
[13] S. A. El-Seoud, "Big data and cloud computing: Trends and challenges," *Int. J. Interact. Mobile Technol.*, vol. 11, no. 2, pp. 1–19, 2017.
[14] P. C. Neves, "Big data in cloud computing: Features and issues," in *Proc. IoTBD*, 2016, pp. 307–314.
[15] B. Matturdi, X. Zhou, S. Li, and F. Lin, "Big data security and privacy: A review," *China Commun.*, vol. 11, no. 14, pp. 135–145, Apr. 2014.
[16] L. kacha and A. Zitouni, "An overview on data security in cloud computing," 2018, *arXiv:1812.09053*. [Online]. Available: http://arxiv.org/abs/1812.09053
[17] D. Dave, "Cloud security issues and challenges," in *Big Data Analytics*. Springer, 2018, pp. 499–514.
[18] A. Narang and D. Gupta, "A review on different security issues and challenges in cloud computing," in *Proc. Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Sep. 2018, pp. 121–125.
[19] B. Seth and S. R. D. Kumar, "Securing bioinformatics cloud for big data: Budding buzzword or a glance of the future," in *Recent Advances in Computational Intelligence*. Cham, Switzerland: Springer, 2019, pp. 121–147.
[20] C. Hasti and A. Hasti, "Data security in cloud-based analytics," in *Big Data Analytics*. Singapore: Springer, 2018, pp. 89–96.
[21] L. Kacha and A. Zitouni, "An overview on data security in cloud computing," in *Proceedings of the Computational Methods in Systems and Software*. Cham, Switzerland: Springer, 2017.
[22] V. N. Inukollu and S. S. R. A. Ravuri, "Security issues associated with big data in cloud computing," *Int. J. Netw. Secur. Appl.*, vol. 6, no. 3, p. 45, 2014.
[23] S. Sabir, "Security issues in cloud computing and their solutions: A review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 11, p. 4, 2018.
[24] P. Srivastava and R. Khan, "A review paper on cloud computing," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 8, no. 6, pp. 17–20, 2018.
[25] O. Harfoushi, B. Alfawwaz, N. A. Ghatasheh, R. Obiedat, M. M. Abu-Faraj, and H. Faris, "Data security issues and challenges in cloud computing: A conceptual analysis and review," *Commun. Netw.*, vol. 6, no. 1, pp. 15–21, 2014.
[26] E. S. A. Ahmed and R. A. Saeed, "A survey of big data cloud computing security," *Int. J. Comput. Sci. Softw. Eng.*, vol. 3, no. 1, p. 7, 2014.
[27] M. B. Sridhar, A. Koushik, and J. Nehru, "A study of big data analytics in clouds with a security perspective," *Int. J. Eng. Res.*, vol. V6, no. 1, pp. 5–9, Dec. 2016.
[28] E. Brynjolfsson, L. M. Hitt, and H. H. Kim, "Strength in numbers: How does data-driven decisionmaking affect firm performance?" Apr. 2011.
[29] I. A. Ajah and H. F. Nweke, "Big data and business analytics: Trends, platforms, success factors and applications," *Big Data Cognit. Comput.*, vol. 3, no. 2, p. 32, Jun. 2019.
[30] M. U. Khan and A. W. Khan, "A security model for big data usage on cloud computing," *Int. J. Comput. Sci. Softw. Eng.*, vol. 3, no. 1, pp. 78–85, 2014.
[31] M. Troester, "Big data meets big data analytics," SAS Inst. Inc, Cary, NC, USA, White Paper 105777_S81514_0512, 2012, pp. 1–13.
[32] P. Ranabhat, "Secure design and development of IoT enabled charging infrastructure for electric vehicle: Using CCS standards for DC fast charging," Bachelor thesis, Dept. Eng., Metropolia Univ. Appl. Sci., Helsinki, Finland, 2018.
[33] S. Sicular. *Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused With Three 'V's*. Accessed: Jan. 2021. [Online]. Available: https://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definitionconsists-of-three-parts-not-to-be-confused-with-three-vs/?sh=15e32e5f42f6
[34] T. A. Patil, S. Pandey, and A. T. Bhole, "A review on contemporary security issues of cloud computing," in *Proc. 1st Int. Conf. Intell. Syst. Inf. Manage. (ICISIM)*, Oct. 2017.
[35] K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, "Big data security and privacy in healthcare: A review," *Procedia Comput. Sci.*, vol. 113, pp. 73–80, Jan. 2017.
[36] A. Kote, P. V. K. Raja, and M. V. Raju, "Cloud data security challenges and its solutions," *IJCCER*, vol. 3, no. 5, pp. 89–92, 2015.
[37] S. Mazumdar, D. Seybold, K. Kritikos, and Y. Verginadis, "A survey on data storage and placement methodologies for cloud-big data ecosystem," *J. Big Data*, vol. 6, no. 1, pp. 1–37, Dec. 2019.
[38] P. Jain, M. Gyanchandani, and N. Khare, "Enhanced secured map reduce layer for big data privacy and security," *J. Big Data*, vol. 6, no. 1, pp. 1–17, Dec. 2019.
[39] L. Wang, Z. Yang, and X. Song, "SHAMC: A secure and highly available database system in multi-cloud environment," *Future Gener. Comput. Syst.*, vol. 105, pp. 873–883, Apr. 2020.
[40] P. Joshi, "Big data gets cloudy: Challenges and opportunities," in *Modern Mathematical Methods and High Performance Computing in Science and Technology*. Singapore: Springer, 2016, pp. 193–206.
[41] Z. Y. Chang, Y. L. Zhao, D. M. Qian, and N. Wang, "Research on cloud computing and network security in digital manufacturing platform," *Appl. Mech. Mater.*, vols. 484–485, pp. 219–222, Jan. 2014.
[42] H. Sohail, "Challenges and opportunities in big data and cloud computing," in *Proc. Int. Conf. Future Intell. Veh. Technol.* Cham, Switzerland: Springer, 2016, pp. 175–181.
[43] A. Adhikari, "Trust issues for big data about high-value manufactured parts," in *Proc. IEEE 2nd Int. Conf. Big Data Secur. Cloud (Big Data Security), IEEE Int. Conf. High Perform. Smart Comput. (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS)*, Apr. 2016, pp. 24–29.
[44] P. D. Pise and N. J. Uke, "Efficient security framework for sensitive data sharing and privacy preserving on big-data and cloud platforms," in *Proc. Int. Conf. Internet things Cloud Comput.*, Mar. 2016.
[45] J. Gao, L. Lei, and S. Yu, "Big data sensing and service: A tutorial," in *Proc. IEEE 1st Int. Conf. Big Data Comput. Service Appl.*, Mar. 2015.
[46] S. S. Kumar and J. S. L. Santhosh, "Big data security issues and challenges in cloud computing environment," *Int. J. Innov. Eng. Technol.*, vol. 6, pp. 297–306, 2015.

[47] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in Software Engineering," Tech. Rep., 2007, pp. 1–65.
[48] A. W. Khan and S. U. Khan, "Critical success factors for offshore software outsourcing contract management from vendors' perspective: An exploratory study using a systematic literature review," *IET Softw.*, vol. 7, no. 6, pp. 327–338, Dec. 2013.
[49] S. U. Khan, M. Niazi, and R. Ahmad, "Barriers in the selection of offshore software development outsourcing vendors: An exploratory study using a systematic literature review," *Inf. Softw. Technol.*, vol. 53, pp. 693–706, Jul. 2011.
[50] W. Chen and S. B. S. Blainey, "A security-as-a-service solution for applications in cloud computing environment," in *Proc. Commun. Netw. Symp.*, 2018, pp. 1–9.
[51] I. A. T. Hashem, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, 2015.
[52] S. Rizvi, J. Ryoo, J. Kissell, and B. Aiken, "A stakeholder-oriented assessment index for cloud security auditing," in *Proc. 9th Int. Conf. Ubiquitous Inf. Manage. Commun.*, Jan. 2015.
[53] R. R. Parmar, S. Roy, D. Bhattacharyya, S. K. Bandyopadhyay, and T.-H. Kim, "Large-scale encryption in the Hadoop environment: Challenges and solutions," *IEEE Access*, vol. 5, pp. 7156–7163, 2017.
[54] A. Luntovskyy and J. Spillner, "Security in distributed systems," in *Architectural Transformations in Network Services and Distributed Systems*. Wiesbaden, Germany: Springer Vieweg, 2017, pp. 247–308, doi: 10.1007/978-3-658-14842-3_7.
[55] M. Bahrami and M. Singhal, "The role of cloud computing architecture in big data," in *Information Granularity, Big Data, and Computational Intelligence*. Springer, 2015, pp. 275–295.
[56] S. Wu and C. Wang, "Big data security framework based on encryption," in *Proc. Int. Conf. Cloud Comput. Secur.* Cham, Switzerland: Springer, 2018, pp. 528–540.
[57] M. Almorsy, J. Grundy, and I. Müller, "An analysis of the cloud computing security problem," 2016, *arXiv:1609.01107*. [Online]. Available: http://arxiv.org/abs/1609.01107
[58] T. Guelzim and M. Obaidat, "Cloud computing systems for smart cities and homes," in *Smart Cities and Homes*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 241–260.
[59] M. Bland, *An Introduction to Medical Statistics*. London, U.K.: Oxford Univ. Press, 2015.
[60] V. M. Deshpande and M. K. A. N. Bihani, "Optimization of security as an enabler for cloud services and applications," in *Cloud Computing for Optimization: Foundations, Applications, and Challenges*. Cham, Switzerland: Springer, 2018, pp. 235–270.
[61] L. Xu and W. Shi, "Security theories and practices for big data," in *Big Data Concepts, Theories, and Applications*. Cham, Switzerland: Springer, 2016, pp. 157–192.
[62] N. A. N. Adnan and S. Ariffin, "Big data security in the web-based cloud storage system using 3D-AES block cipher cryptography algorithm," in *Proc. Int. Conf. Soft Comput. Data Sci.* Singapore: Springer, 2018, pp. 309–321.
[63] R. Barona and E. A. M. Anita, "A survey on data breach challenges in cloud computing security: Issues and threats," in *Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT)*, Apr. 2017.
[64] R. Buyya, "A manifesto for future generation cloud computing: Research directions for the next decade," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–38, 2018.
[65] K. Gai, Y. Wu, L. Zhu, and M. Qiu, "Privacy-preserving data synchronization using tensor-based fully homomorphic encryption," in *Proc. 17th IEEE Int. Conf. Trust, Secur. Privacy Comput. Communications/ 12th IEEE Int. Conf. Big Data Sci. Eng. (TrustCom/BigDataSE)*, Aug. 2018, pp. 1149–1156.
[66] R. Vora and K. P. G. Raval, "An era of big data on cloud computing services as utility: 360° of review, challenges and unsolved exploration problems," in *Proc. 1st Int. Conf. Inf. Commun. Technol. Intell. Syst.* Cham, Switzerland: Springer, vol. 2, 2016, pp. 575–583.
[67] S. Pandey and G. U. M. P. Munshi, "Data security in cloud-based applications," in *Data Science Landscape*. Springer, 2018, pp. 321–326.
[68] C. Stergiou and K. E. Psannis, "Algorithms for big data in advanced communication systems and cloud computing," in *Proc. IEEE 19th Conf. Bus. Informat. (CBI)*, Jul. 2017, pp. 196–201.
[69] C. L. Stergiou, "Secure machine learning scenario from big data in cloud computing via internet of things network," in *Handbook of Computer Networks and Cyber Security*. Cham, Switzerland: Springer, 2020, pp. 525–554.
[70] C. Stergiou, K. E. Psannis, B. B. Gupta, and Y. Ishibashi, "Security, privacy & efficiency of sustainable cloud computing for big data & IoT," *Sustain. Comput., Informat. Syst.*, vol. 19, pp. 174–184, Sep. 2018.
[71] E. Mendes, C. Wohlin, K. Felizardo, and M. Kalinowski, "When to update systematic literature reviews in software engineering," *J. Syst. Softw.*, vol. 167, pp. 174–184, Sep. 2020, Art. no. 110607.
[72] M. Akram, A. N. Al-Kenani, and J. C. R. Alcantud, "Group decision-making based on the VIKOR method with trapezoidal bipolar fuzzy information," *Symmetry*, vol. 11, no. 10, p. 1313, Oct. 2019.
[73] P. Chatterjee and S. Chakraborty, "A comparative analysis of VIKOR method and its variants," *Decis. Sci. Lett.*, vol. 5, no. 4, pp. 469–486, 2016.
[74] M. Das, X. Tao, and J. C. P. Cheng, "BIM security: A critical review and recommendations using encryption strategy and blockchain," *Autom. Construction*, vol. 126, Jun. 2021, Art. no. 103682.
[75] B. A. Y. Alqaralleh, T. Vaiyapuri, V. S. Parvathy, D. Gupta, A. Khanna, and K. Shankar, "Blockchain-assisted secure image transmission and diagnosis model on internet of medical things environment," *Pers. Ubiquitous Comput.*, pp. 1–11, Feb. 2021.
[76] M. M. A. Baig, "Cloud computing ethical issues: A review paper to investigate and provide suggestions for solving data privacy issues of cloud computing," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 7, pp. 1433–1438, 2021.
[77] D. Fu, S. Hu, L. Zhang, S. He, and J. Qiu, "An intelligent cloud computing of trunk logistics alliance based on blockchain and big data," *J. Supercomput.*, pp. 1–16, May 2021.
[78] J. Zhang, *A Systematic Literature Review of Data Privacy Issues in Cloud BI*, vol. 77, no. 8. Auckland, New Zealand: Auckland University of Technology, 2021.
[79] S. Juyal, S. Sharma, and A. Shankar Shukla, "Security and privacy issues in unified IoT-based skin monitoring system," *Mater. Today, Proc.*, vol. 46, Feb. 2021.
[80] S. Riaz, A. H. Khan, M. Haroon, S. Latif, and S. Bhatti, "Big data security and privacy: Current challenges and future research perspective in cloud environment," in *Proc. Int. Conf. Inf. Manage. Technol. (ICIMTech)*. IEEE, 2020, pp. 977–982.
[81] K. J. Modi, P. D. Shah, and Z. Prajapati, "Security and privacy in big data computing: Concepts, techniques, and research challenges," in *Quantum Cryptography and the Future of Cyber Security*, Hershey, PA, USA: IGI Global, 2020, pp. 236–256.
[82] W. Aljedaibi and A. A. Alsulami, "A CMMI-based method for software development process assessment," *Applying CMMI Process Saudi Arabia*, 2021.
[83] M. A. Akbar, A. A. Khan, A. W. Khan, and S. Mahmood, "Requirement change management challenges in GSD: An analytical hierarchy process approach," *J. Softw., Evol. Process*, vol. 32, no. 7, pp. 1–31, Jul. 2020.
[84] N. Zarbakhshnia, Y. Wu, K. Govindan, and H. Soleimani, "A novel hybrid multiple attribute decision-making approach for outsourcing sustainable reverse logistics," *J. Cleaner Prod.*, vol. 242, Jan. 2020, Art. no. 118461.
[85] M. Shameem, A. A. Khan, M. G. Hasan, and M. A. Akbar, "Analytic hierarchy process based prioritisation and taxonomy of success factors for scaling agile methods in global software development," *IET Softw.*, vol. 14, no. 4, pp. 389–401, Aug. 2020.
[86] G. Wieland and H. Zeiner, "A survey on criteria for smart home systems with integration into the analytic hierarchy process," in *Proc. Int. Conf. Decis. Support Syst. Technol.* Cham, Switzerland: Springer, 2021, pp. 55–66.
[87] A. Angitha and M. Supriya, "Ranking of educational institutions based on user priorities using AHP-PROMETHEE approach," in *Advances in Computing and Network Communications*. Singapore: Springer, 2021, pp. 127–142.
[88] C. Xu, "Using AHP-entropy approach to investigate the key factors on FinTech service," *J. Comput.*, vol. 32, no. 1, pp. 200–211, 2021.
[89] W. Liu, J. Pang, S. Yang, N. Li, Q. Du, D. Sun, and F. Liu, "Research on security assessment based on big data and multi-entity profile," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, vol. 5, Mar. 2021, pp. 2028–2036.

ABUDUL WAHID KHAN received the Ph.D. degree in computer science from the University of Malakand. He is currently working as an Assistant Professor with UST Bannu, Pakistan.

MASEEH ULLAH KHAN received the B.S.C.S. degree from COMSATS University Islamabad. He is currently a M.S.C.S. Scholar with the University of Science and Technology Bannu, Khyber Pakhtunkhwa, Pakistan. He is performing his services at the Education Department, Bannu, Khyber Pakhtunkhwa.

JAVED ALI KHAN received the B.Sc. degree in computer software engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2009, the M.Sc. degree in software engineering from Baheria University Islamabad, Pakistan, in 2013, and the Ph.D. degree in software engineering from Tsinghua University, China, in 2021. He has been working as an Assistant Professor with the Department of Software Engineering, University of Science and Technology Bannu, Pakistan, since December 2011. He has published more than ten papers in reputable journals and conferences in requirements engineering. His areas of interests include requirements engineering, CrowdRE, argumentation and argument mining, feedback analysis, empirical software engineering, sentiment analysis, and opinion mining.

ARSHAD AHMAD received the M.S. degree in software engineering from Blekinge Tekniska Högskola (BTH), Sweden, in 2008, and the Ph.D. degree in computer science and technology (specialization in software engineering) from Beijing Institute of Technology, China, in 2018. From 2010 to 2014, he worked as a Research Assistant at the Fraunhofer Institute of Experimental Software Engineering (IESE), Germany, and Vienna University of Technology, Austria, respectively. He is currently working as an Assistant Professor of software engineering and computer science with the Department of IT and CS, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur, Pakistan. He has published several research papers in well reputed peer-reviewed international journals and conferences. His current research interests include requirements engineering, text mining, opinion mining, sentiment analysis and machine learning, and among others.

KHALIL KHAN received the B.S. degree in electrical engineering and the M.S. degree in computer engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2007 and 2012, respectively, and the Ph.D. degree from the Signals and Communication Laboratory, University of Brescia, Italy, in 2016. He is currently working as an Assistant Professor with the Department of Information Technology and Compute Science, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Pakistan. His research interests include a wide range of topics within digital image processing, machine learning, and computer vision.

MUHAMMAD ZAMIR received the Ph.D. degree in mathematics from the University of Malakand. He is currently working as an Assistant Professor with UST Bannu, Pakistan.

WONJOON KIM received the Ph.D. degree in industrial engineering from Seoul National University, in 2017. He is currently an Assistant Professor with the Division of Future Convergence (HCI Science Major), Dongduk Women's University. His research interests include human factor, artificial intelligence, deep learning, and process of new product development with particular interests in product design and improvement.

MUHAMMAD FAZAL IJAZ (Member, IEEE) received the B.Eng. degree in industrial engineering and management from the University of the Punjab, Lahore, Pakistan, in 2011, and the Dr.Eng. degree in industrial and systems engineering from Dongguk University, Seoul, South Korea, in 2019. From 2019 to 2020, he worked as an Assistant Professor with the Department of Industrial and Systems Engineering, Dongguk University, Seoul. He is currently working as an Assistant Professor with the Department of Intelligent Mechatronics Engineering, Sejong University, Seoul. He has published numerous research articles in several international peer-reviewed journals, including IEEE Access, *Sensors*, *Symmetry*, *Journal of Food Engineering*, *Applied Sciences*, *Asia Pacific Journal of Marketing and Logistics*, and *Sustainability*. His research interests include machine learning, blockchain, healthcare engineering, the Internet of Things, supply chain management, big data, and data mining.