REVIEW ON DATA MANAGEMENT IN CLOUD COMPUTING

MUHSIN MUSTAFA HASSAN[1*], ASADULLAH SHAH[2]

[12]Kulliyyah of Information and Communication Technology (KICT), International Islamic University Malaysia (IIUM), Gombak, Malaysia

**Corresponding author: mohsn.mustafa@gmail.com*

ABSTRACT: The present cloud computing time opens another road for organizations to consider moving their information data management solutions to the cloud. It is demonstrated through the recent successful deployments that data management applications are one of the most suitable candidates for deployment in the cloud. Data management in cloud computing brings many challenges as well as advantages. This paper represents the current state of data management solutions in the cloud and the current direction of research in data management in cloud computing. It is necessary to contemplate that the cloud-based data management is a very fluid field, which means technologies are changing rapidly. It is also important that the industry is investing more as compared to the other field of data management, thus making an interesting subject for detailed study.

KEY WORDS: *Cloud, Data Management, Cloud Storage, Data Security, Data Privacy*

1. INTRODUCTION

The start of last decade clearly presented that the pace and volume of data being generated is exceeding the current capacity of 'institutions' data management. Cloud-based data management is in turn helping to realize the potential of large-scale data management solutions by giving effective scaling of resources. In a cloud-based data management scenario, institutions rent storage and computing power to make the data management applications work rather than making considerable capital in-house investment for infrastructure.

Cloud computing is gaining a high popularity in the IT industry as it offers significant benefit by freeing them from the low-level tasks of setting up basic hardware and software infrastructures, and thus enabling them to focus on innovation and strengthening the business value for their services by providing benefaction of reduced cost and easy and fast deployment of resources. It is a paradigm shift into the next generation data centers hosted by large infrastructure companies such as Amazon, Google, Yahoo, Microsoft, or Sun.

Cloud computing is an extremely successful paradigm of service-oriented computing and has revolutionized the way computing infrastructure is abstracted and used. The three most popular cloud paradigms include: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). The concept however can also be extended to Database as a Service or Storage as a Service. Elasticity, pay-per use, low upfront investment, low time to market, and transfer of risks are some of the major enabling features that make cloud computing a universal paradigm for deploying novel applications which were not

economically feasible in a traditional enterprise infrastructure setting. Scalable database management systems, both for update intensive application workloads, as well as decision support systems—are thus a critical part of the cloud infrastructure.

The main motivation of this paper is to study and analyze 10 of the most recent research papers in data management in cloud computing, to study all suggested data quality dimensions, compare them and find any possible gaps to be filled by future studies.

2. METHODOLOGY

The methodology used in this research is systematic literacy review (SLR). The stages in this research are divided into 4 (four) main parts, namely: stage 1, stage 2, stage 3, and stage 4 as used by Buettner and Buettner (2016). The research question in this study was: ‘‘What are the Data Management issues in cloud computing?”. The solution used to answer this research question is the Systematic Literature Review (SLR) method by conducting a paper selection process that fits the selected criteria (Al-Ruithe et al., 2019). With the SLR method, we can find out the trends of research topics that are of great interest to previous researchers, so that this can be used as a reference for further research. The importance of the SLR method in this research is that researchers can identify and analyze journals systematically according to the recommended stages related to journals about the issues and components of data management in cloud computing.

The process of selecting this article was carried out in stages, namely:

- Stage 1: identify from the source database based on the search criteria, at this stage the number of articles that can be identified is 96 articles.
- Stage 2: identify the selected articles in the first stage based on the article titles relevant to the research topic, at this stage 32 articles can be identified.
- Stage 3: selection of articles based on the suitability of abstract content and keywords, resulting in 16 articles.
- Stage 4: this is the final stage of removing several articles that are not eligible based on the search criteria, at this stage 10 articles can be identified. The distribution of the number of articles from 2010 to 2020 can be seen in Fig. 1 below.

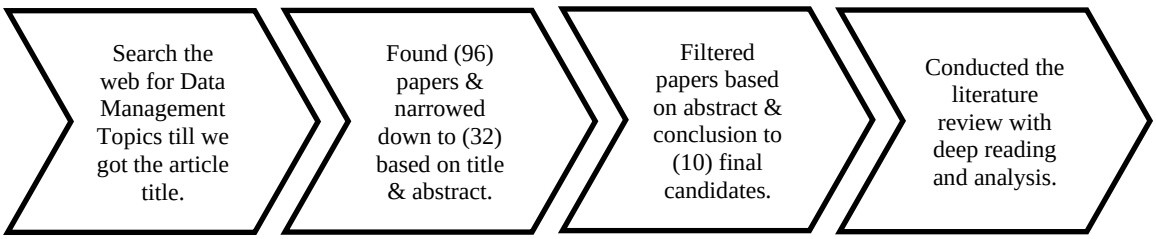

Fig. 1. Methodology followed for the literature review

3. LITERATURE REVIEW

Data storage through cloud can save costs because cloud service providers already provide the infrastructure so that it can reduce the provision of their own infrastructure (Ren et al., 2012). Several other reasons for companies migrating to the cloud are because: (1) cloud computing services have scalability, which means

that they can meet the needs of information technology resources according to the company needs; (2) the cloud provider has provided settings for both hardware configuration and software updates or server settings and others, so that companies as the cloud service users are more focused on developing better innovative products; and (3) the cloud provider has a data center that provides fast and efficient computing services, so this will have an effect on high performance in the cloud compared to the data center owned by the company.

Based on the data from cisco.com, it is estimated that in 2020 the cloud data centers will process by 93%, while in 2021 the workload of data centers will increase to 94%. The factor driving the migration of workloads from the traditional data centers to the cloud data centers is a greater degree of virtualization in the cloud space, which enables dynamic deployment of workloads in the cloud-to-cloud service demands.

3.1. SERVICES AND CHARACTERISTICS

The current widely accepted cloud computing definition is based on the NIST definition that identifies five essential cloud characteristics (P. Mell and T. Grance):

- On-demand self-service.
- Broad network access and diversity of client devices.
- Resource pooling that allows providers to serve multi-tenant customers by managing resource utilization more efficiently using virtualization, resource partitioning and workload balancing.
- Rapid elasticity that allows scaling resources dynamically.
- Measured service with the pay-per-use business model.

3.2. ISSUES IN CLOUD COMPUTING

Behl (2011) explores cloud computing security issues and highlights the key research challenges that include:

- Availability and Performance — this issue can be resolved through well-formed SLA (service level agreement) coined with real-time monitoring.
- Malicious Insiders — the cloud service providers cannot restrict their employees, contractors and other trusted people who have access to the secure data of customers through supply chain management.
- Outside Attacks — for example, the hackers can get access to the data. To resolve this issue, the network perimeter should be protected through firewalls.
- Service Disruptions — it can occur when no more resources are available for other customers, and this may cause customer dissatisfaction. This issue can be resolved by ensuring that the connections are coming from known IP pool and DNS.

3.3. CLOUD DATABASE STORAGE MANAGEMENT

Because of the ubiquitous nature of cloud computing, there is a rapid growth in the number of applications which leverage various cloud platforms, resulting in a tremendous increase in the scale of the data generated as well as consumed by such applications. Scalable database management systems (DBMS) — both for update intensive application workloads, as well as decision support systems — are thus a critical part of the cloud infrastructure. Conventional DBMS deals with structured data which is held in databases along with its metadata, while cloud databases can be used for unstructured, semi-structured data or structured data. The data stored in files of various types where the metadata was either unavailable or incomplete is called unstructured data. Cloud databases are able to support in changing the storage requirements of Internet-savvy users who deal more with unstructured data, user created content such as documents and photos.

As shown in Fig. 2 below, a data management life cycle is depicted to utilize large scale data. The first step is to collect digital data in the field, followed by data organization and annotation, which requires sophisticated computer infrastructure to streamline the workflow and continuous stream of collected data. A system which fascinates cloud computing has a large amount of data, which can be collected by using a database and a data warehouse for analysis. The data must be transferred between two spatially separated, tightly coupled locations that is front end and back end. To bridge this gap, we need a mobile data management system equipped with abstracted program and data hosting infrastructure and means of data access.

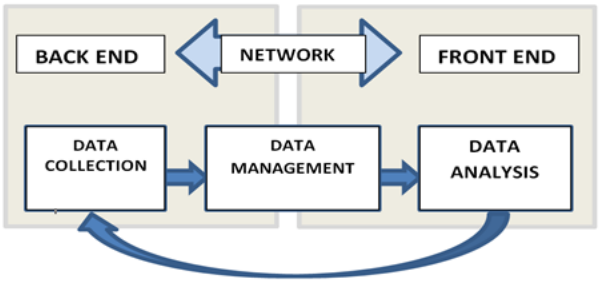

Fig 2. Data management cycle

3.4. List of literatures

Adaptive query execution for data management in the cloud (Popescu, Dash, Kantere and Ailamaki, 2010)

IoT Data Management using Cloud Computing and Big Data technologies (Gupta & Godavarti, 2020)

Intelligent data management and security in cloud computing (Ogiela, Ogiela and Ko, 2020)

Secured Cloud computing for data management using big data for small and medium educational institutions (Jain, 2020).

A review on cloud computing and data management in cloud (Jangra and Bhatia, 2014)

Big data and cloud computing: Trends and challenges (El-Seoud, El-Sofany, Abdel-fattah, Mohammed, 2017)

Cloud computing challenges and opportunities: A survey (Moghaddam et al., 2015).

Data management in cloud computing (Vankunde, 2018)

Data storage and data management in cloud computing (Josh and Joshi, 2015)

Cloud computing: The future of big data Management (Ouf & Nasr, 2015)

3.5. Research Questions and Objectives

The objective of Popescu, Dash, Kantere and Ailamaki (2010) was to architect adaptive software that efficiently processes both long-running and short-running queries in the same storage layer. The proposed model introduces the adaptive query execution scheme that could take in the query and probability of node failures to decide where the query can be executed in either parallel DBMS or MapReduce based on the cost model.

Gupta & Godavarti (2020) designed a system that manages big data created by IoT devices via Rest APIs. The system is synchronized with scalability to handle huge volume data, cost effectiveness to compute resources and provide efficient storage as well as speed to increase the rate at which the data is processed. The proposed system claims to provide an excellent platform for big data analysis using cloud computing services.

The issue of intelligent data management and security in cloud computing is explained by Ogiela, Ogiela and Ko (2020), where they have elaborated the new approaches for the utilization of linguistic methods for data management and secure sharing protocols in hierarchical structures. In their article, they have elaborated graph and formalism that can be utilized in division and secret management of information in a secured manner using data sharing techniques. This paper provided a unique approach to fabricate universal division protocols that works in multiple hierarchical structures.

Jain (2020) proposed a security approach and methods that can be utilized in cloud computing for data management to be applied in small and medium educational institutions. In his model, he proposed securing and storing data in the cloud using an encryption technique in frequency domain using EDDDWT. To retrieve the original data from the cloud, decryption is applied where the data management system converts frequency domain to spatial or time domain. The proposed solution is aimed to improve efficiency and safety in educational institutions by preserving privacy and offering network and data security.

In their article, Jangra and Bhatia (2014) reviewed on the overall concept of cloud computing, cloud database and ways to provide database as a service in a network. They further discussed the concepts on cloud model, cloud storage techniques and deployment models. In addition, the paper highlighted the risk factors associated with cloud data storage with its solutions.

El-Seoud, El-Sofany, Abdel-fattah, Mohammed (2017) introduced the characteristics, models, analytics, and changes of big data. They also discussed big data management systems (MapReduce and Hadoop) that satisfy the processing of big data distributed in computing environment. Furthermore, this paper has drawn out the advantages, risks and challenges that may result from the big data and cloud computing integration.

Moghaddam et al. (2015) conducted a survey featuring researchers and service providers to point out the challenges and the important issues concerning cloud-base environment. For each challenge, the most appropriate and relevant solution was drawn. This paper aimed to provide a broad explanation of the cloud computing visions and reveal the areas which require further research.

Vankunde (2018) investigated the present state of data management solutions in the cloud with reference to data management architecture, data privacy and data security in the cloud. This paper also pointed out the current preferences to research that should be prioritized concerning data management in cloud computing. The paper claimed to provide an insight into the past and current cloud-based data management and research preferences.

Josh and Joshi (2015) modified their own data model to provide security to application stored in public cloud by using security techniques encryption, message digest, hash function and decryption. This model had the motive of maintaining integrity, privacy and security of data stored of in public cloud.

Ouf & Nasr (2015) analyzed the adoption process in managing enterprise's big data in situations of high complex environment especially the uncertainty about external and internal ad technological factors surrounding the adoption of new technology. The paper proposed that the usage of cloud computing for managing data enterprises improves the management of big data, improves performance, and makes the data scalable and available.

3.6. Research Methodologies

Research methodology describes the path at which the researcher has used to conduct their research to accomplish the set objectives (Sileyew, 2019). Based on the chosen 10 articles previously described, there are four types of methodology that were adopted to describe them.

a) Review methodology: In this method, the findings of the previous literature are broadly integrated to collect data, to facilitate theory development and to highlight areas where the topic of research is desperate (Synder, 2019). The articles which have adopted this methodology are:
   - A review on cloud computing and data management in cloud (Jangra and Bhatia, 2014)
   - Data management in cloud computing (Vankunde, 2018)

b) Qualitative methodology: It is a method that explores and dig deep on the real world problems. This type of method normally gives a description of the characteristics of the research problem to gain deeper understanding of the problem. The articles which utilized this method are:
   - Big data and cloud computing: Trends and challenges (El-Seoud, El-Sofany, Abdel-fattah, Mohammed, 2017)
   - Intelligent data management and security in cloud computing (Ogiela, Ogiela and Ko, 2020)
   - Data storage and data management in cloud computing (Josh and Joshi, 2015)
   - Cloud computing: The future of big data Management (Ouf & Nasr, 2015)

c) Survey method: This method involves gathering data and drawing conclusions from a variety of respondents. The article which opted this research method is:
   - Cloud computing challenges and opportunities: A survey (Moghaddam et al., 2015).

d) Experimental design method: In this method, a model or experiment is carried out to test the hypothesis formulated to solve the problem in hand. The article which opted this method are:
   - IoT Data Management using Cloud Computing and Big Data technologies (Gupta & Godavarti, 2020)

- Adaptive query execution for data management in the cloud (Popescu, Dash, Kantere and Ailamaki, 2010)
- Secured Cloud computing for data management using big data for small and medium educational institutions (Jain, 2020).

3.7. Research Results

The main motive of this paper is to identify the area of improvement of data management in cloud computing. 10 articles have been selected covering the topic of cloud computing. The results and gaps identified in each of the articles have been summarized in the following table to identify the desperate area in cloud computing that requires extensive research.

Research paper and author	Research results	Research gaps
Adaptive query execution for data management in the cloud (Popescu, Dash, Kantere and Ailamaki, 2010)	The adaptive query execution scheme solves the problem of short-running query in which its performance is affected on long running query analysis such as MapReduce. The scheme provides two order magnitude speed-ups for short running queries as compared to Map reduce while still preserves the benefits of long running queries.	The algorithm of adaptive query execution scheme is probabilistic in nature. It is therefore prone to make mistakes and wrong choices when the aggregated failure rate is close to the running time of query. Osprey system which implements faults in MapReduce for distributed DBMS is not designed for short running queries which increases the cost of checkpoints.
IoT Data Management using Cloud Computing and Big Data technologies (Gupta & Godavarti, 2020)	A secured, fault tolerant, cost effective and scalable system capable of handling the fast growing and huge variety of data from IoT devices was developed.	The system was designed to gather data via IoT devices only. The system should be extended to provide data management from the worldwide internet. The system still needs to be improved by implementing more fault tolerant, scalable and security mechanism.
Secured Cloud computing for data management using big data for small and medium educational institutions (Jain, 2020).	The encryption and compression technique used to provide security to data and network in the cloud for small and medium educational institution was found to be efficient and accurate as compared to the existing system.	The designed system was limited to securing cloud data for only small and medium enterprise in educational institutions. There is a need for other institution such as health centers to try out the system to secure their cloud data.
Intelligent data management and security in cloud computing (Ogiela, Ogiela and Ko, 2020)	The new approaches of linguistic method for the utilization for data management and security in hierarchical structures were proposed to as layered and stricter approaches. The graph structure shows the greatest possibilities of application in, among others, the area of cloud computing.	The graph structure was not performed on parallel architecture. The intelligent data management in cloud computing was not designed to secure data in the form of CAPTCHA. This area requires improvement.
A review on cloud computing and data management in cloud (Jangra and Bhatia, 2014)	The concept of cloud computing from cloud model, deployment models and storage techniques were discussed in detail. The	The solutions to the risks were discussed in brief with not enough explanation.

	authors highlighted that the risk factors associated with cloud computing are minimum availability and poor performance, outside attacks, malicious insiders, and services disruptions, all of which are related to security issues.	The authors also identified the need to have a mobile data management system equipped with abstract program, data hosting infrastructure and means of data access to transfer data from front end and back end using database and data warehouse for analysis in cloud database storage management.
Big data and cloud computing: Trends and challenges (El-Seoud, El-Sofany, Abdel-fattah, Mohammed, 2017)	In this paper the unique idea was the benefits and the risks associated with the integration of big data and cloud computing. The identified advantages of the integration are the usage of cloud computing to store big data, improve their performance in data storage and process power availability, and will make big data provisioning faster, easy, and feasible. The identified challenge and risk is the security vulnerability in deploying big data on cloud environment. Another challenge is cloud environment may not have locations to accommodate big data because of its nature to be in different locations. This will make the data processing very difficult.	The authors did not highlight the specific solutions to the risks and challenges of big data and cloud computing integration which calls on the need for further research.
Cloud computing challenges and opportunities: A survey (Moghaddam et al., 2015).	After the survey was carried out, the authors identified the challenges of cloud computing and have categorized in 6 major groups which are security and privacy, data management and resource allocations, load balancing, interoperability and communication between cloud, migration to clouds and compatibility and scalability and availability which still limit the utilization of cloud computing. The authors identified that the security and privacy are the major challenge. The solution suggested to this challenge was the application of cryptography algorithm to improve resistance of possible attacks and unpredicted events.	The solutions to overcome the load balancing problems were not highlighted, which demands further research.
Data management in cloud computing (Vankunde, 2018)	Security is still a major problem in cloud storage especially to organization utilizing public cloud. The solution proposed is the use of cryptographic cloud storage which offers private storage with security of public cloud and with cost savings of public cloud.	The paper only covered the parts which are most discussed by most researchers, which are data privacy, data management architecture and data security in cloud. Areas such as scalability which include storage of petabyte data and online query processing were not discussed.

	In privacy, there is a limit of disclosure of confidential data. The solutions provided are data perturbation method and data swapping. Another problem with privacy is legislation and government aspect of data storage which demand access to the data stored in every computer. In terms of data architecture, hybrid cloud featuring the properties of public and private cloud is proposed to reduce the problem of security in cloud computing.	The hybrid model discussed in the data management architecture is still challenged by the security issues especially in public cloud. The strategies to effectively manage data in hybrid cloud still needs further exploration. The issue of reduction in data quality during grouping, merging, and swapping of data in the cloud also demands an extensive study.
Data storage and data management in cloud computing (Josh and Joshi, 2015)	The simple model to provide the data security to the data stored in public cloud was designed and proven to successfully maintain privacy and integrity of the data stored in the cloud. The authors also mentioned that the reasons for the achievement of the cloud computing paradigm for cloud data management model are the usage of product hardware in large scale to develop the economies of scale, elasticity and pay as go model of payment.	The model still requires improvement in the design of flexible and scalable system that can provide data management and data security as service in cloud computing.
Cloud computing: The future of big data Management (Ouf & Nasr, 2015)	The authors proposed a cloud database for the efficient management of big data in cloud computing. This technology allows enterprises to store relational data on cloud servers in service provider data centers. The relational data model allows the storage of data in the system even in the case of network failures.	Only the schematic model of cloud database was resented lacking the systematic model and architecture details. The paper did not highlight how the cloud database is going to increase the effectiveness and efficiency of Big Data management in enterprises.

4. CONCLUSION

Cloud computing makes a better use of distributed resources, combine them to achieve higher throughput and be able to solve large scale computation problems. Cloud computing provides a simultaneous way to lower costs, increase responsiveness and flexibility, and improve quality of service. Google, Amazon, Yahoo, and other Internet service providers, as well as IBM, Microsoft and other IT companies have proposed their own cloud computing strategy, while various telecom operators paid great attention to cloud computing. In this paper the basic concepts like different cloud models, deployment models and cloud storage techniques are discussed.

REFERENCES

A. Behl, ‖Emerging Security Challenges in Cloud Computing‖, IEEE international Conference Information and Communication Technologies (WICT), 2011. Pp.1 to 4

El-Seoud, S., El-Sofany, H., Abdelfattah, M., & Mohamed, R. (2017). Big Data and Cloud Computing: Trends and Challenges. *International Journal of Interactive Mobile Technologies (Ijim)*, *11*(2), 34. doi: 10.3991/ijim. v11i2.6561

Gupta, S., & Godavarti, R. (2020). IoT Data Management Using Cloud Computing and Big Data Technologies. *International Journal of Software Innovation*, *8*(4), 50-58. doi: 10.4018/ijsi.2020100104

Jain, N. (2020). SECURED CLOUD COMPUTING FOR DATA MANAGEMENT USING BIG DATA FOR SMALL AND MEDIUM EDUCATIONAL INSTITUTIONS. *International Journal of Computer Engineering & Technology (IJCET)*, *11*(2), 21-30.

Jangra, A., & Bhatia, V. (2014). A Review on Cloud Computing and Data Management in Cloud. *IJITKM*, (ISSN 0973-4414), 27-32.

Josh, K., & Joshi, J. (2015). DATA STORAGE SECURITY AND DATA MANAGEMENT IN CLOUD COMPUTING. *GLOBAL JOURNAL FOR RESEARCH ANALYSIS*, *5*(2), 47-48. doi: 10.36106/gjra

Moghaddam, F., Ahmadi, M., Sarvari, S., Eslami, M., & Golkar, A. (2015). Cloud computing challenges and opportunities: A survey. *2015 1St International Conference on Telematics and Future Generation Networks (TAFGEN)*. doi: 10.1109/tafgen.2015.7289571

Ogiela, L., Ogiela, M., & Ko, H. (2020). Intelligent Data Management and Security in Cloud Computing. *Sensors*, *20*(12), 3458. doi: 10.3390/s20123458

Ouf, S., & Nasr, M. (2015). Cloud Computing. *International Journal of Cloud Applications and Computing*, *5*(2), 53-61. doi: 10.4018/ijcac.2015040104

P. Mell and T. Grance, —The NIST Definition of Cloud Computing, ‖ National Institute of Standards and Technology Special Publication 800-145, 2011

Popescu, A., Dash, D., Kantere, V., & Ailamaki, A. (2010). Adaptive query execution for data management in the cloud. *Proceedings Of the Second International Workshop on Cloud Data Management - Clouddb '10*. doi: 10.1145/1871929.1871933

Sileyew, K. (2020). Research Design and Methodology. *Cyberspace*. doi: 10.5772/intechopen.85731

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal Of Business Research*, *104*, 333-339. doi: 10.1016/j.jbusres.2019.07.039

Vankundre, A. (2018). Data Management in the Cloud Computing. *International Journal for Scientific Research & Development*, *5*(12), 227-230.