Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Key challenges and research direction in cloud storage

Mr. Nikhil S. Gajjam, Dr. T. Gunasekhar

Computer Science& Engineering Department, K L Deemed to be University Guntur 522502, India

ARTICLE INFO

Article history:
Received 29 December 2020
Accepted 19 January 2021
Available online xxxx

Keywords:
Cloud Computing
Cloud database
Availability
Cloud Storage
Consistency

ABSTRACT

Cloud computing has gained the importance in recent time because of its unique characteristics like elasticity, resilience in delivering compute, storage, network, and resource management in form of utility services over the Internet. Cloud computing area has attracted attention and interest due to the ever-increasing demand for reliable and cost-effective infrastructure and service delivery. This paper primarily focuses on the survey of issues and open challenges of cloud database in cloud computing storage. The aim of this survey is to provide a perspective to a researcher a general overview of issues, open challenges, and future trends in cloud storage with primarily focuses on cloud databases. The paper also tabulates comparison of different methods and approaches of researchers in tackling issues pertaining to data availability through data replication, data partitioning, data management, and data placement under the context of cloud enabled technologies.
© 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

1. Introduction

Cloud Computing is a successful computing paradigm based on grid computing, and has gained the importance in last decade. Cloud Computing is a process where services are provided over the internet. According to NIST definition, Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to computing resources like network, servers, storage, application and services that can be acquired or released with minimum management efforts or service provider information [1]. A cloud is a place which provides on demand access to information technology (IT) and computing resources which include virtualized hardware, operating systems, network, storage, databases, and applications [2] Fig. 1 Table 1 Table 2.

When it comes to data storage, Most of the Cloud Service Provider has multiple regions distributed around the world. Region is nothing but the physical location where multiple Data centers are present. Each region has physically separated Availability zones. All these Availability zones are connected with high speed network. Each and every Availability zone contains one or more data centers.

Cloud storage is a collection of storage resources which are virtually distributed globally across the different data centers. Storage includes blocks, files, objects, relational databases (includes SQL and NoSQL). Cloud service provider provides storage as a service with characteristics like consistency, scalable storage, recovery management, multi-tenancy. Advantages of deploying data intensive applications on cloud are reduced time, reduced cost, reduced operational cost, unlimited throughput [3]. When it comes to Database-as-a-Service (DBaaS) then it must support availability, ACID properties, transaction support, elasticity etc. As database is key part of many applications, research is going on in increasing performance of the cloud hosted database applications. Availability, Scalability, Elasticity and performance are the major goal of cloud hosted databases. Different challenges of Cloud Hosted Databases are True Elasticity, Data Replication, Consistency, Live Migration, SLA Management, Transaction Support, and Benchmarking [4].

Fig. 2.a) represents overall classification of cloud storage system. Cloud storage categorized into 2 parts Unstructured and Structures. Unstructured storage includes Block storage, File Storage and Object Storage. Structured storage includes RDBMS and NOSQL.

Block Storage: It is used as primary or secondary storage. Data is stored in fixed size blocks. Network latency is low so it used for the high-performance application. Generally, it is used for storing Operating system or databases. Scalability of block storage is low as compared to File and Object type of cloud storage

File Storage: It is generally used for unstructured data or secondary storage. An advantage of a file storage system is in a shared

https://doi.org/10.1016/j.matpr.2021.01.609
2214-7853/© 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

Please cite this article as: Mr. Nikhil S. Gajjam and Dr. T. Gunasekhar, Key challenges and research direction in cloud storage, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2021.01.609

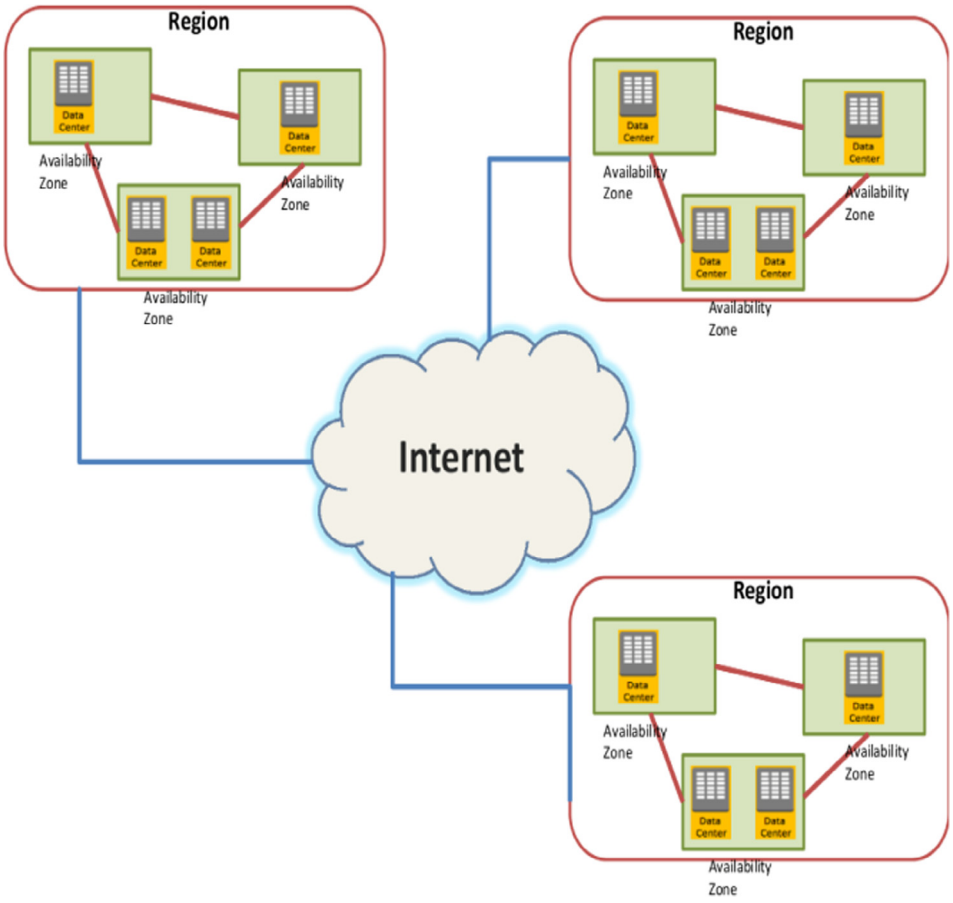

Fig 1. Region and Availability Zones.

Table 1
Difference in Block, File and Object storage Scheme.

Parameter	Block	File	Object
Applicable	OS/ structured data	File backup	Unstructured data
Scalability	Low	Moderate	High
Protocols	NFS/ CIFS	iSCSI / iSER	HTTP
Performance	Very High	Moderate	Low

Table 2
Difference in SQL and NoSQL.

	Relational Database	NOSQL
Performance	High	Low
Reliability	Low	High
Scalability	High	High but complex
Storage	High	Medium
Consistency	Low	High

file system where many chunks mount on the single volume. File storage is relatively slow as compared to other storage schemes.

Object Storage: It is chunks of data objects accessed by the unique identifier. It is best suited for unstructured data. Advantages of this scheme are platform independent. These are accessed over the internet so HTTP protocol is required. Scalability is very high compared to Block and File.

RDBMS: Relational database is the collections of structured data, tables where each table is the collections of rows. Structured Query Language (SQL) is used for accessing, retrieving and updating the relational database. Indexing is the most important factor in relational database. Advantages of relational database are categorizing, storing the data, simplicity as well as it allows user to use complex query because of structured data. Some of the concerned factors of the Relational database are scalability, security and performance. But at present also relational database take upper hand over NoSQL because of its ACID properties. Amazon

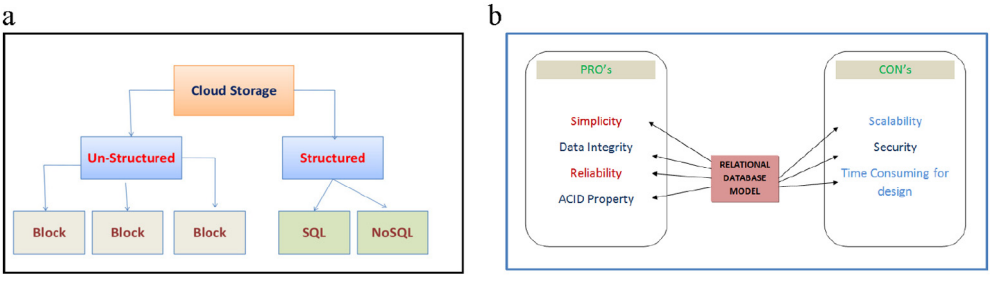

Fig. 2. a) Types of cloud storage b) Pros and Cons of Relational Database System.

RDS and Microsoft SQL Azure are the famous examples of the relational database on a cloud.

NoSQL: The applications which rely on large storage and without ACID properties generally use NoSQL databases. Data stored in NoSQL database is in one of the form of document, key-value, graph, object oriented or column based. NoSQL databases allows user to access huge data easily, quickly.

There are 5 major types of NoSQL databases. Key-value store database, Column-oriented database, Document store database, graph database, Object-oriented database.

Key-value based: In key value based database, data is stored in a key-value pair where values are indexed using a unique key. Key-value based data model is simple [5] as it used hash table as an index. Examples of a key-value based database are Amazon DynamoDB, RIAK, and Cassandra etc.

Column-Oriented based: Column-oriented databases a table for storage but keeps column together. All columns are treated separately. The database is designed for storing data as a section of columns of data rather than rows. Examples are HBase, Google BigTable. The advantage of these types of databases is searching is very fast.

Document-based: In a document-based database, each document is identified by a unique key. Each document is stored as the JSON object. For document retrieval, only the document identifier and ranges of these identities can be used, so the documents can't be searched by their contents directly [6]. Best suited for document or file storage. Flexibility, scalability, and performance are the advantages of the document based database. Examples are MongoDB, Raven DB, CouchDB etc.

Graph-based: In this type of database, relations are represented as a graph and all elements are interconnected with each other. Examples are Neo4J, Polyglot, and TITAN.

Object Oriented: Object-oriented database is a combination of object-oriented programming and database. Application with complex object relationships uses this type of storage [5].

2. Issues in cloud storage

Researchers have worked on various issues related challenges of Cloud storage are Data management, True Elasticity, Data Replication, Consistency, Live Migration, SLA Management, security etc.

2.1. Data management

Data management of data stored at multiple geo-distributed data centers is an important issue. Different challenges of data management on the federated cloud are parallelism, recovery, consistency in replication, reliability [7]. Data management as a service focuses on the design of a scalable and elastic data management system. Three primary concerns [8] for any application includes levels of data consistency, high availability, and partition tolerance. Sherif Sakret al. [9] presents a survey of number of perspectives for deploying data-intensive applications in the cloud. Various issues have been identified like scalability, consistency, cost for processing data in the paper. MapReduce technique has been used for query processing of structured data [10] where input query is divided into different subqueries and replicas are mapped to k + 1 subqueries.

Divyakant Agrawal et al. [11] carried a research work on Managing Geo-replicated Data in Multi-datacenters. Providing full transactional support to replicated data across the different data center is needed. This paper describes two algorithms one for optimistic concurrency control and other for providing distributed transaction across datacenters.

Gang Chen et al. [12] presents a paper on Federation in Cloud Data Management: Challenges and Opportunities. The author introduces a federation framework to manage heterogeneous clouds along with different challenges and solution. Asymmetric Hardware Capability, Dynamic Resource Sharing, Reliability, Consistency in Data Replication are the different federation challenges in cloud data management.

DYFRAM [13], is a dynamic approach based on access pattern to tables for fragmentation and allocation in distributed database systems. Fragmentation, replication, and reallocation in his approach is based on recent access history. SQLMR [14] is a hybrid approach which combines SQL-based data processing and MapReduce data processing. SQLMR is useful in increasing performance and scalability as compared to SQL and NoSQL data processing. Reddy, B. T et al. [38] represented a technique of deduplication of cloud data to keep only one copy of the file to achieve greater management of the data. This paper also addresses on the techniques to provide secure deduplication of data. Genetic algorithm based approach can also be used to prevent data deduplication in cloud [50].

2.2. Data partitioning

The major objective of the partitioning is to increase the performance of the database for input work. Two major partitioning techniques are horizontal and vertical partitioning. Horizontal partitioned data is easier to manage in terms of operations like backup and restore. Fragmentation is also useful in load balancing. Load balancing can be achieved if fragmented data is distributed among servers based on the capacity of the server. Fragmentation plays important role in reducing network traffic. If fragments are placed closer to the application then those fragments can answer to the smaller queries. Partition is done for majorly linear speedup and Improving availability.

Two partitioning techniques are horizontal and vertical. Mostly used horizontal partitioning techniques are round robin and hash partitioning. Many strategies consider only to partition and distribution of data objects. Performance and scalability are the two major concerns when used in widely distributed systems [15]. Yunqi Ye et al. [15] presents the paper on Cloud Storage Design Based on Hybrid of Replication and Data Partitioning. This approach provides better security, performance and response time. DYFRAM [13], is a dynamic approach based on access pattern to tables for fragmentation and distribution in distributed database systems. Fragmentation, replication, and reallocation in this approach is based on recent access history. Rimma Nehme et al. [16] proposed automatic Partitioning of the data in Parallel Database. The author describes technique for finding the best partitioning structure in the distributed environments. Outcome of the technique is tables should be replicated and distributed according to a specific column(s) for minimizing cost for evaluating similar work.

Optimal data partitioning in cloud computing system with random server assignment [17] is another data partitioning policy that can effectively mitigate effects of the co-resident attacks through minimizing user's losses.

Elasca [18] is Workload-Aware Elastic and Scalable technique for Distributed databases. In this technique a workload aware optimizer algorithm identifies the optimal partitioned data and its placement. And this technique performs 79% more efficiently when data transfer cost is considered. Demonstration of this technique has been performed on OLTP databases.

2.3. Data placement

Data placement is the key factor which impacts on performance in the cloud environment. Geographically Distant Data Sharing, Cli-

ent Mobility, Data Inter-dependencies are some issues in data placement.

Volley is a log based approach for data placement [20] for Geo-Distributed Cloud Services. This algorithm is based on data access patterns and client locations. The author carried out the experiments on Microsoft's Live Messenger and Live Mesh for automatic data placement at distributed data centers.

Intelligent Database Placement [21] approach helps DBaaS providers to achieve efficient resource allocation, minimize the disturbance to the system due database migration, and maximize Cloud resource utilization. An author wants to improve the results by considering resource consumption.

SWORD [22] is another methodology for data placement and replica selection based on workload. Advantages of the approach are minimizing the consumption of various computing resources.

2.4. Availability through replication

Process of sharing information to provide availability is called as replication. Consistency between redundant resources is major issue in replication to improve reliability, fault-tolerance, or accessibility. Availability and efficiency of communication resources plays major role in deciding performance of cloud applications. Data resources can be replicated closer to the application to reduce network latency. Generally replication is needed to provide for fault tolerance and availability for the application. Bandwidth consumption and users response time are the key factors in providing replication to the application. Systems that provide both reliability and availability are often said to be fault-tolerant. The main objective of data replication is to provide availability and to reduce response time and total bandwidth consumption. Advantages of replication: Less energy consumption, network bandwidth, and communication delay, improves reliability, fault tolerance and accessibility. Challenges of replication: Data selection, Consistency in replicated data, sophisticated management, load balancing and recovery.

Different models for Efficient Data Replication in Cloud Computing Datacenters [23] have been presented a model on efficient data replication strategy for geo-distributed data center cloud with energy consumption and bandwidth factor taken into consideration. Paper concluded that replicated data should place closer to data consumer to reduce energy consumption and bandwidth.

RepliC [24] is the new technique for database replication in cloud environment with features like quality of service, elasticity, and multi-tenancy. Full replication strategy is used in RepliC technique. Decision support system [25] is used for replication in distributed database. Cristina L. Abad et al. [26] proposed method DARE, a distributed adaptive data replication algorithm aims at improving data locality. Probabilistic sampling and a competitive aging algorithm is used in this approach for finding number of replicas to allocate for each file and placement of those replicas. Yunqi Ye, Liangliang Xiao [27] present paper on Designing a cloud storage based on replication and partitioning. The author proposed two-level DHT (TDHT) approach for widely distributed cloud storage to address performance and scalability problems. Security, dependability and performance requirements of cloud storage are the different issues solved by the research work.

Minimizing the replication cost and maximize the expected availability of objects are the two major issues addressed by Yaser Mansouri et al. [28].This proposed approach is useful in determining the optimal location of the replicas for objects. To determine the proper locations of data replicas based on varying query load or disasters, Nicolas Bonvin et al. [29] presented an approach Skute. A Light-weight Data Replication for Cloud Data Centers Environment [30] strategy decides dynamically the number of replicas and location of data nodes for replication for improving availability. Another threshold-based dynamic data replication strategy [31] has been proposed and is based on Fast Spread. This strategy has improved the performance in terms of total response time and total bandwidth consumption. Log sentiment analysis and intelligent replication can be used for improving availability [32]. Advantages of the research are reducing cost, optimize resources and availability.

Dynamic Data Replication Strategy [33] has been proposed to Increase System Availability in Cloud Computing Environments. This paper consists of survey of different replication strategy for distributed computing environments. Author has also given relationship between availability and number of replicas. Another important aspect is identifying the popular data. This paper presents a dynamic replication algorithm which based on evaluating popular replicas and replica placement when popularity of data crosses certain benchmark.

DYFRAM [13], is a dynamic approach based on access pattern to tables for fragmentation and allocation in distributed database systems. Fragmentation, replication, and reallocation in this approach is based on recent access history.

Quanlu Zhang et al. [35] presents a new approach CHARM. It works in two phases, first is the selection of suitable cloud and redundancy strategy for minimizing cost and increasing availability, the second one is the distribution of data based on data access pattern. Tung Nguyen et al.[36] presents Differentiated Replication Strategy in Data Centers. This replication strategy is based on user requirements and system capability. This technique can be further improved by considering performance and durability factors. Bo Mao et al.[37] presents a paper on "Exploiting Workload Characteristics and Service Diversity to Improve the Availability of Cloud Storage Systems". The workload characteristics and the diversity of cloud providers can used to improve the cloud storage availability in Cloud-of-Clouds. This approach has advantageous in terms of better I/O performance and cost.

2.5. Data security

When it comes to data in cloud environment, many researchers have identified various issues like data theft, data unavailability, data privacy, data integrity etc. [42].

From the consumers perspective, data security and privacy is another important aspect while availing cloud services [39]. Deyan Chen et al. [39] presented a paper on Data Security and Privacy Protection Issues in Cloud Computing. This paper consists of data security and privacy protection issues, its solution and future directions.

R. Velumadhava Rao et al. [40] published a paper titled Data Security Challenges and Its Solutions in Cloud Computing. This paper highlights various security challenges in cloud. Exploring Data Security Issues and Solutions in Cloud Computing [41] is another article which focused on data security issues and its solutions.

Vurukonda, N. et al. [43] presented an approach for achieving flexible and scalable access control for the data in cloud computing. It is a new decentralized grained access control approach. Security and privacy concerns are important issues in Cloud data migration[44]. Ravindranadh, K. et al. [44] presented a honey-encryption based cryptographic methodology for outsourced data.

Dhote, B. L. et al. [45] published a paper on Trust and Security to Shared Data in Cloud Computing: Open Issues. This paper focused on the data security from data owner, users and cloud provider. A Modified revocable and searchable ABE technique [46] is a another technique of security for the stored data, by encrypting it using secrete key generated by attribute set and both TPA and Cloud.

Prasad, G. S. et al. [47] proposed system to provide the integrated security for data in cloud storage. Gowtham Kumar, N.

et al.[48] proposed work plan is to use encryption algorithms to eliminate concerns about data privacy in order to enhance security in the cloud M. Ranjeeth Kumar et al. [49] proposed a new approach which overcomes the limitation of Transmitted Team Key Management (TTKM) for file sharing system on cloud. Apart from these there are many factor on which researchers worked like load balancing, energy efficiency [54], resource scheduling, networking, SLA, etc [56].

Researcher	Data Management	Replication	Data Partitioning	Consistency	Data Placement	Key Parameters/ Usefulness	
[7]	+	+	–	+	–	• Large data management	
[9]	+	–	–	+	–	• Data Intensiveapplications	
[10]	+	–	+	–	–	• Scalability • Efficiency	
[11]	+	+	–	+	–	• Georeplicated DBMS	
[12]	+	+	–	+	–	• HeterogeneousCloud management	Data
[13]	+	+	+	–	–	• Distributeddatabase • ReducedCost	
[14]	+	–	–	–	–	• Scalability • Performance	
[15]	–	+	+	–	–	• Security • Performance	
[16]	–	+	+	–	–	• Minimizes Cost	
[17]	–	–	+	–	–	• Elasticity • Scalability	
[19]	–	–	+	–	–	• Scalability • Suitedfor OLTP database	
[20]	–	–	–	–	+	• Location aware placement	
[21]	–	–	–	–	+	• Efficient resource utilization	
[22]	–	+	–	–	+	• Minimum resource Utilization	
[24]	–	+	–	–	–	• Multi- tenancy • Elasticity • Qos	
[25]	–	+	–	–	+	• Fault tolerance	
[26]	–	+	–	–	+	• Performance • Achieve better locality	
[27]	–	+	+	–	–	• Performance • Scalability	
[28]	–	+	–	–	+	• Cost effective	
[29]	–	+	–	–	–	• Key-value storage	
[30]	–	+	–	–	+	• Dynamic • File storage system	
[31]	–	+	–	–	–	• Performance	
[32]	–	+	–	–	–	• Cost • Performance	
[33]	–	+	–	–	–	• Reduces response time	
[34]	–	+	+	–	–	• Improve scalability	
[35]	–	+	–	–	–	• Cost • Multi-cloud environment	
[36]	–	+	–	–	–	• Resource utilization	
[37]	–	+	–	–	–	• Better I/O performance • Cost efficiency	
[38]	+	–	–	–	–	• security	

3. Open research challenges and future trends

Multicloud: Multicloud is the use of two or more public cloud service provider for utilizing the services of Platform-as-a-Service and Software-as-a-Service. Multicloud is preferred in some situations as it prevents downtime, data security, cost-effective. Use of multicloud provides better security, ease of replication and fragmentation of application data and logic. Application management is one of the major problems in multi-cloud environment.

DevOps in Cloud: DevOps has emerged as a new technology that bridges the gap between developers and operation team. DevOps tools can manage and configure the resources like storage, networking and computing. So efficient deployment is possible with DevOps tools in Cloud Computing environments. Another advantage of using DevOps is speed and productivity. Integration of cloud computing and DevOps allows developers and operations work as a single unit in single common language.

Machine Learning in Cloud: Recently many of the researchers have done their research on parallelization of machine learning

and data mining problems in a distributed environment. Software-as-a Service provider provides a solution to many machine learning problems through RESTful APIs and interfaces [5158].

Big data in cloud: Though Cloud Computing is widely accepted in most of the IT companies and organization but when it comes to big data in cloud computing still many research challenges are there which includes scalability, availability, data integrity, data transformation, data quality, data heterogeneity, privacy and legal issues[52]. This paper presents a classification, a conceptual view, and a cloud services model for big data. Storing, processing, data management of big data in cloud computing environment are the areas still to be discussed.

IOT in cloud and Fog Computing: Cloud of Things is an integrated area of cloud computing and IOT. New opportunities in IOT include complex analysis, data mining and real-time processing can be solved using Cloud of things. IOT is hundreds of devices connected with lots of data gathering and processing. This can be done efficiently in cloud environment due to its unique characteristics. Recently Fog computing is the new area emerged and located at the edge of the network. Yi, Shanhe et al. [53] presented new research directions in Fog computing like networking, Quality of Service, Interfacing, Programming Model, Computation Offloading, Resource Management, Security and Privacy.

Research Topic	Related Factors	Open Challenge
Data Storage	Fragmentation	-Providing workload aware dynamic fragmentation technique for reducing data migration
	Data Placement	-Providing a proper data placement strategy by considering factors like prediction of resource consumption and data locality
	Data management	-Design of data management strategy with consideration of buffer management and memory cost
	Replication	-Designing an efficient and dynamic data replication strategy with consistency and fragmentation strategy-Designing an efficient data replication and consistent management strategy in geo distributed cloud by considering data pattern and network load factors.

4. Conclusion

This article mainly focuses on different issues in cloud storage like data management, data partitioning, replication, data placement along with different terminologies in storage. Most of the research carried out by taking factors like optimizing monetary cost, reducing network latency, optimal storage, increasing performance, optimal resource utilization are taken into consideration. Choosing correct methodology and strategy depends on characteristics and requirement of the application.

Multiple data storage and management strategies have been discussed in this article which may useful in improving the performance of Applications based on Database-as-a-Service (DBaaS) in the cloud.

CRediT authorship contribution statement

Nikhil S. Gajjam: Conceptualization, Methodology, Software, Visualization, Writing - Original draft. **T. Gunasekhar:** Data curation, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.
[2] R.N. Calheiros, R. Ranjan, R. Buyya, in: Virtual machine provisioning based on analytical performance and QoS in cloud computing environments, IEEE, 2011, pp. 295–304.
[3] Sakr, S. (2011). a Liu, DM Batista, and M. Alomari, A Survey of Large Scale Data Management Approaches in Cloud Environments, Commun. Surv. Tutorials.
[4] Y. Mansouri, A.N. Toosi, R. Buyya, Data storage management in cloud environments: Taxonomy, survey, and future directions, ACM Computing Surveys (CSUR) 50 (6) (2017) 1–51.
[5] A. Nayak, A. Poriya, D. Poojary, Type of NOSQL databases and its comparison with relational databases, Int. J. Appl. Inform. Syst. 5 (4) (2013) 16–19.
[6] Moniruzzaman, A. B. M., & Hossain, S. A. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. arXiv preprint arXiv:1307.0191.
[7] D.J. Abadi, Data management in the cloud: Limitations and opportunities, IEEE Data Eng. Bull. 32 (1) (2009) 3–12.
[8] V. Gudivada, D. Rao, V. Raghavan, Renaissance in data management systems: SQL, NoSQL, and NewSQL, Comput. (2014).
[9] Sakr, S. (2011). a Liu, DM Batista, and M. Alomari, A Survey of Large Scale Data Management Approaches in Cloud Environments, Commun. Surv. Tutorials.
[10] J. Zhao, X. Hu, X. Meng, October). ESQP: an efficient SQL query processing for cloud data management, in: In Proceedings of the second international workshop on Cloud data management, 2010, pp. 1–8.
[11] D. Agrawal, A. El Abbadi, H.A. Mahmoud, F. Nawab, K. Salem, Managing geo-replicated data in multi-datacenters, Springer, Berlin, Heidelberg, 2013, pp. 23–43.
[12] G. Chen, H.V. Jagadish, D. Jiang, D. Maier, B.C. Ooi, K.L. Tan, W.C. Tan, Federation in cloud data management: Challenges and opportunities, IEEE Trans. Knowl. Data Eng. 26 (7) (2014) 1670–1678.
[13] J.O. Hauglid, N.H. Ryeng, K. Nørvåg, DYFRAM: dynamic fragmentation and replica management in distributed database systems, Distributed Parallel Databases 28 (2–3) (2010) 157–185.
[14] M.J. Hsieh, C.R. Chang, L.Y. Ho, J.J. Wu, P. Liu, SQLMR: A scalable database management system for cloud computing, IEEE, 2011, pp. 315–324.
[15] Y. Ye, L. Xiao, I.L. Yen, F. Bastani, Cloud storage design based on hybrid of replication and data partitioning, IEEE, 2010, pp. 415–422.
[16] Nehme, R., & Bruno, N. (2011, June). Automated partitioning design in parallel database systems. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (pp. 1137-1148).
[17] G. Levitin, L. Xing, Y. Dai, Optimal data partitioning in cloud computing system with random server assignment, Future Generation Computer Systems 70 (2017) 17–25.
[18] T. Rafiq, Elasca: Workload-aware elastic scalability for partition based database systems, Master's thesis University of Waterloo, 2013.
[19] J.M.M. Kamal, M. Murshed, R. Buyya, in: Workload-aware incremental repartitioning of shared-nothing distributed databases for scalable cloud applications, IEEE, 2014, pp. 213–222.
[20] Agarwal, S., Dunagan, J., Jain, N., Saroiu, S., Wolman, A., & Bhogan, H. (2010). Volley: Automated data placement for geo-distributed cloud services.
[21] T. Yu, J. Qiu, B. Reinwald, L. Zhi, Q. Wang, N. Wang, Intelligent database placement in cloud environment, IEEE, 2012, pp. 544–551.
[22] K.A. Kumar, A. Quamar, A. Deshpande, S. Khuller, SWORD: workload-aware data placement and replica selection for cloud data management systems, VLDB J. 23 (6) (2014) 845–870.
[23] D. Boru, D. Kliazovich, F. Granelli, P. Bouvry, A.Y. Zomaya, Models for efficient data replication in cloud computing data centers, IEEE, 2015, pp. 6056–6061.

[24] F.R. Sousa, J.C. Machado, Towards elastic multi-tenant database replication with quality of service, IEEE, 2012, pp. 168–175.
[25] P. Elango, Data replication for the distributed database using decision support systems, Int. J. Comput. Appl. 69 (3) (2013).
[26] Abad, C. L., Lu, Y., & Campbell, R. H. (2011, September). DARE: Adaptive data replication for efficient cluster scheduling. In 2011 IEEE international conference on cluster computing (pp. 159-168). Ieee.
[27] Y. Ye, L. Xiao, I.L. Yen, F. Bastani, Cloud storage design based on hybrid of replication and data partitioning, IEEE, 2010, pp. 415–422.
[28] Y. Mansouri, A.N. Toosi, R. Buyya, Brokering algorithms for optimizing the availability and cost of cloud storage services, IEEE, 2013, pp. 581–589.
[29] N. Bonvin, T.G. Papaioannou, K. Aberer, A self-organized, fault-tolerant and scalable replication scheme for cloud storage, in: In Proceedings of the 1st ACM symposium on Cloud computing, 2010, pp. 205–216.
[30] M.K. Hussein, M.H. Mousa, A light-weight data replication for cloud data centers environment, Int. J. Eng. Innovative Technol. 1 (6) (2012) 169–175.
[31] M. Bsoul, A. Al-Khasawneh, Y. Kilani, I. Obeidat, A threshold-based dynamic data replication strategy, J. Supercomput. 60 (3) (2012) 301–310.
[32] López, C., Heinsen, R., & Huh, E. N. (2017, January). Improving Availability Applying Intelligent Replication in Federated Cloud Storage Based on Log Analysis. In Proceedings of the 2017 International Conference on Machine Learning and Soft Computing (pp. 148-153).
[33] D.W. Sun, G.R. Chang, S. Gao, L.Z. Jin, X.W. Wang, Modeling a dynamic data replication strategy to increase system availability in cloud computing environments, J. Comput. Sci. Technol. 27 (2) (2012) 256–272.
[34] Curino, C., Jones, E. P. C., Zhang, Y., & Madden, S. R. (2010). Schism: a workload-driven approach to database replication and partitioning.
[35] Q. Zhang, S. Li, Z. Li, Y. Xing, Z. Yang, Y. Dai, CHARM: A cost-efficient multi-cloud data hosting scheme with high availability, IEEE Trans. Cloud Comput. 3 (3) (2015) 372–386.
[36] T. Nguyen, A. Cutway, W. Shi, in: Differentiated replication strategy in data centers, Springer, Berlin, Heidelberg, 2010, pp. 277–288.
[37] B. Mao, S. Wu, H. Jiang, Exploiting workload characteristics and service diversity to improve the availability of cloud storage systems, IEEE Trans. Parallel Distrib. Syst. 27 (7) (2015) 2010–2021.
[38] Tirapathi Reddy, B., & Chandra Sekhara Rao, M. V. P. (2017). Data Deduplication In Cloud Storage Using Dynamic Perfect Hash Functions. Journal of Advanced Research in Dynamical and Control Systems, 9(12), 2121-2132.
[39] Chen, D., & Zhao, H. (2012, March). Data security and privacy protection issues in cloud computing. In 2012 International Conference on Computer Science and Electronics Engineering (Vol. 1, pp. 647-651). IEEE.
[40] R.V. Rao, K. Selvamani, Data security challenges and its solutions in cloud computing, Procedia Comput. Sci. 48 (2015) 204–209.
[41] P.R. Kumar, P.H. Raj, P. Jelciana, Exploring data security issues and solutions in cloud computing, Procedia Comput. Sci. 125 (2018) 691–697.
[42] B.T. Rao, A study on data storage security issues in cloud computing, Procedia Comput. Sci. 92 (2016) 128–135.
[43] K. Kamakshaiah, K.V. Rao, M. Subrahmanyam, SABE: Efficient and scalable-filtered access control in distributed cloud data storage, in: Smart Computing and Informatics, Springer, Singapore, 2018, pp. 31–42.
[44] K. Ravindranadh, M.S. Kiran, B.D.S.P. Kumar, D. Priyanka, Data migration in cloud computing using honey encryption, Int. J. Eng. Technol. 7 (2.8) (2018) 230–234.
[45] B.L. Dhote, G.K. Mohan, in: Trust and Security to Shared Data in Cloud Computing: Open Issues, Springer, Singapore, 2019, pp. 117–126.
[46] P.M. Mane, C.M. SheelaRani, High data availability with effective data integrity and user revocation using abe scheme for cloud storage, Int. J. Innovat. Technol. Explor. Eng. 8 (4S2) (2019) 109–113.
[47] G.S. Prasad, D.L. Praneetha, S. Srivalli, B.V. Sukesh, Information security in cloud by using enhanced triple-DES encryption algorithm, Int. J. Innovat. Technol. Explor. Eng. 8 (5) (2019) 679–682.
[48] N. Gowtham Kumar, N. Polala, D. Aruna Kumari, New approach for securing cloud applications, in: Paper presented at the Proceedings of the 2nd International Conference on Inventive Systems and Control, 2018, https://doi.org/10.1109/ICISC.2018.8398886, Retrievedfrom www.scopus.com.
[49] M.R. Kumar, N. Srinivasu, L.C. Reddy, Dynamic and secure authentication using IKC for distributed cloud environment, Springer, Singapore, 2018, pp. 107–117.
[50] PanduRanga Rao, K. V., Krishna Reddy, V., &Yakoob, S. K. (2018). Dynamic secure deduplication in cloud using genetic programming 10.1007/978-981-10-3223-3_48.
[51] [54] Pop, Daniel. Machine learning and cloud computing: Survey of distributed and saas solutions. arXivpreprint arXiv:1603.08767 (2016).
[52] Hashem, Ibrahim Abaker Targio, et al. "Therise of "big data" on cloud computing: Review and open research issues. Information systems 47 (2015): 98-115.
[53] Shanhe Yi, Cheng Li, Qun Li, A survey of fog computing: concepts, applications and issues, Proceedings of the 2015 workshop on mobile big data, 2015.
[54] V.P. Dendukuri, U.L. Soundharya, G.K. Chaitanya, T. GunaShekar, Energy efficient approach to reduce carbón emission: The dark side of data center, J. Adv. Res. Dynam. Contr. Syst. 10 (7) (2018) 1729–1734.
[55] N.S. Dey, T. Gunasekhar, A comprehensive survey of load balancing strategies using hadoop queue scheduling and virtual machine migration, IEEE Access 7 (2019) 92259–92284.
[56] S.P. Praveen, K.T. Rao, B. Janakiramaiah, Effective allocation of resources and task scheduling in cloud environment using social group optimization, Arab. J. Sci. Eng. 43 (8) (2018) 4265–4272.