A REVIEW ON QUALITY OF SERVICE AWARE ALGORITHM FOR DATA REPLICATION ACROSS DATA STORES

[1]Dr. S. Sujiya, [2]Mr. S. Rajesh, [3]Mr. E. Srinivasan.

[1]Assistant Professor, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[2]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[3]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

Abstract: - *The IT systems sent to satisfy their business functionalities and to provide good Quality of Service (QoS) parameters such as availability, scalability and performance. The failures are normal rather than exceptional in case of cloud computing. To improve system availability enhancing the system performance is an important factor. Recently the number of cloud storage users has increased abundantly. The reason is that the Cloud Storage system reduces the burden of maintenance. It provides high availability, reliability and also it is most suitable for high volume of data storage. In order to provide high availability and reliability, the systems introduce redundancy. The Data Replication is rendering little bit threat about the Cloud Storage System for the users and for the providers it is a big challenge to provide efficient Data Storage. Thus, this paper analyzed the various existing data replication strategies and pointed out the major issues affected by the Data Replication.*

***Keywords**: - Cloud, Quality of Service (QoS), Algorithm, Data Replication, Data Store.*

1. INTRODUCTION

The cloud computing is a new concept for delivering computing service and largely satisfies emerging requirements of the information technology. It has claimed itself to be of great benefit for users and organizations because it can dramatically reduce the expenses and provide an aid to manage information technology systems without any hassle [1]. Due to its large number of advantages, Cloud has been increasingly adopted in many areas, such as banking, e-commerce, retail industry, and academics etc.

Cloud computing can be defined as a specialized distributed computing model, which is dynamically configured and delivered on demand.. The NIST provides a formal definition of the cloud computing as follows: “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.

Cloud storage services have its advantages and disadvantages. The main advantages of storage services are capital cost savings, because users do not need to invest their money to own storage servers nor do they have to maintain these servers, and scalable; since users can easily increase or decrease their storage capacity based on their needs [2]. In addition, more features are added to cloud storage services such as file sharing and synchronization which make it more appealing for users to use. However, the main disadvantages that pushes users away from adopting cloud storage is usually whether it is sufficiently secure or not.

This paper is organized as: chapter 2 explains the basics of cloud computing and its basics in the cloud model. Moreover, it explains the cloud storage services in more detailed and provides a detailed analysis about data model. Chapter 3 explains various brokering algorithms and its characteristics. Finally, in chapter 5, we will conclude the paper and list some directions for future work.

2. CLOUD COMPUTING SERVICES & CHARACTERISTICS

Cloud computing is a distributed computational model over a large pool of shared-virtualized computing resources (e.g., storage, processing power, memory, applications, services, and network bandwidth), where customers are provisioned and de-provisioned recourses as they need. Cloud computing represents a vision of providing computing services as public utilities like water and electricity. The architecture of cloud computing can be split in two: front-end and back-end [3]. The front-end represents cloud customers, organizations, or applications (e.g., web browsers) that use the cloud services. The back-end is a huge network of data centers with many different applications, system programs, and data storage systems. A conceptual framework of cloud computing architecture is illustrated with its two main parts.

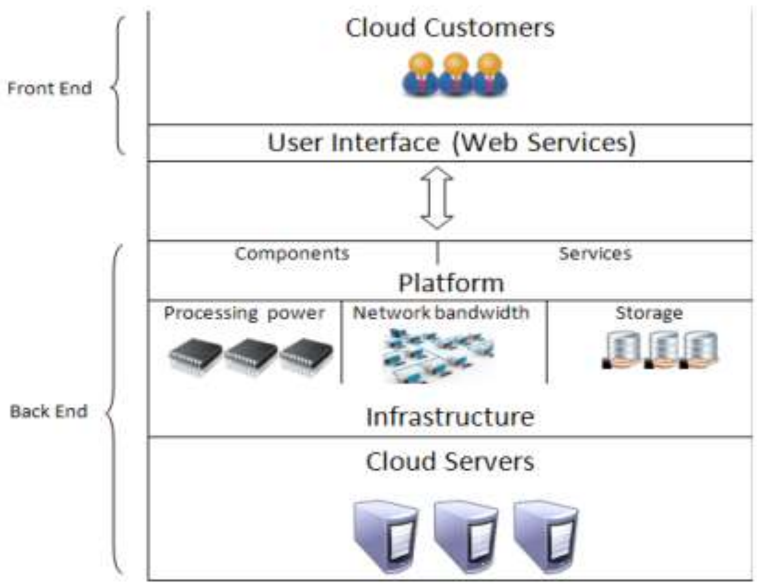

Fig 1: - Conceptual Cloud Computing Framework Architecture

Cloud computing services can be categorized into:

- Application-as-a-Service (AaaS).
- Platform-as-a-Service (PaaS).
- Infrastructure-as-a-Service (IaaS).

The widely used model of cloud computing services is the AaaS model, in which the customers have access to the applications running on the cloud provider's infrastructure. Google Docs, Google Calendar, and Zoho Writer are known examples of this model. In the PaaS model, the customers can deploy their applications on the provider's infrastructure under condition that these applications are created using tools supported by the provider [4]. The cloud service provider (CSP) hosts a set of software and development tools on its servers to be used by the developers to create their own applications.

The cloud computing architecture can be deployed under different models:

- Public cloud.
- Private cloud.
- Hybrid cloud.

2.1. Cloud Characteristics

To better understand the core concepts and technologies in the cloud, we extract from the NIST with five attributes. These attributes describe a cloud based system as a general model providing metered on demand services to his clients. These characteristics are presented as follows:

- On-Demand Self-Service.
- Broad Network Access.
- Shared Resources.
- Elasticity.
- Measured Service.

One of the main principles of Cloud Computing is the `as-a-Service' paradigm in which some services are offered by a Cloud Service Provider (CSP) to customers for use [5]. These offered services are often categorized using the SPI Service Model. This model represents the different layers/levels of service that can be offered to users by cloud service providers over the different application domains and types of cloud available. Clouds can be used to provide as-a-Service: software to use, a platform to develop on, or an infrastructure to utilize.

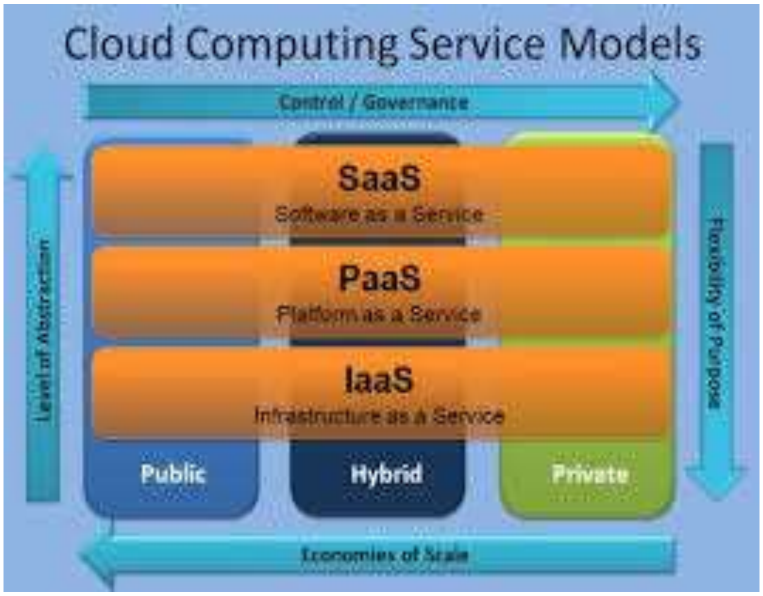

Fig 2: - Cloud Computing Service Model

2.2. Infrastructure as a Service (IaaS)

Infrastructure as a Service (IaaS) is a service that can provide the functionalities of a whole infrastructure including storage, networks, any platform and any number of desktops. The customers can make use of this service by configuring a virtual machine on the infrastructure, on which an operating system is installed.

2.3. Platform as a Service

In the Platform as a Service (PaaS) model, the CSP offers a development platform on top of the services delivered with IaaS. The CSP offers a development platform, on which applications can be built. In other words, software developers can develop their application through virtual development platform, accessible via a Web browser, without the need to install the software building tools on their own computer. This helps the developers to later distribute or deploy their apps to the cloud easily. In order to avoid confusion of this service with SaaS, it is good to imagine it as a cloud OS. The providers of the service enable its users to install their applications on a platform, which can provide any operating system or even emulate various types of hardware [6].

2.4. Software as a Service

SaaS is a very popular service in which cloud service providers deliver software applications over the Web. A SaaS provider deploys their software, which is hosted on their own server infrastructure or use another vendor's hardware, on user's demand .This operation is usually done using a licensing model where applications may be licensed directly to an organization, group of users or, a user or, or through a third party that manages multiple licenses between user organizations, such as an ASP. The user then can be able to access the applications through any well defined and Internet device, which is most probably a Web browser.

3. CLOUD DATA REPLICATION

The main apprehension for the users storing the data in the cloud is to preserve the data and recover it whenever required. Any server failure should not result in data loss. Cloud applications include gaming, voice and video conferencing, online office, storage, backup, social networking. The performance of these applications depends largely on the availability of high performance communication resources and network efficiency [7] [8]. Data replication is a commonly used technique to increase the data availability. It requires a high bandwidth data throughput path. Cloud replicates the data and stores them strategically on multiple servers located at various geographic locations. Replication ensures consistency; improves availability and reliability by creating multiple copies of the same data on different storage devices and geographical locations. Replication and availability play a major role in fault tolerance.

"Data replication" is an appropriate technique to manage data files. Data replication is to create multiple copies of data in multiple sources in order to reduce access time and bandwidth consumption. It also guarantees data reliability and load balancing for the system.

Data is distributed across the cloud. This has to be made available to the applications that want to use it. The performance must not be degraded. The data access speed should be increased, keeping the load balanced in the system. Scalability and availability are the two major factors to improve the performance of the cloud. Creating replicas is also one of the important strategies to achieve the above. Replication also reduces access latency and bandwidth consumption.

3.1. Data Replication

Replication is creating multiple copies of an existing entity [9]. Replication increases availability of resources. It also provides consistency and reliability by creating multiple copies of the same data on different sites. Replication also provides minimum access cost, shared bandwidth utilization and delay time by replicating data. The value of replication is to provide transparent, flawless access to resources in the event of a system failure.

Replication can be extended across a computer network so that storage devices can be located in physically separated facilities. Users access nearby replicas and increase the throughput in case of failure to maintain the transmission of data. There are advantages of storing the data at more than one site. If a server with the required data fails, a system can operate using replicated data. This concept maintains availability. The data is stored at multiple sites. The requested data is fetched from the nearest source from where the request originated. This increases the performance of the system. The benefits of replication do not come without overheads of creating, maintaining and updating the replicas. Replication can greatly improve the performance [10]. There is a performance overhead in replication technique as it takes time to recover data from other sites and restart the service again. The advantage is fault can be tolerated and availability can be increased.

3.2. Challenges in Data Replication

- Data Consistency: Maintaining data integrity and consistency in a replicated environment is most important. High precision applications may require strict consistency of the updates made by transactions.
- Downtime during new replica creation: If strict data consistency is to be maintained, performance is severely affected if a new replica is to be created. Sites will not be able to fulfill request due to consistency requirements.

- Maintenance overhead: If the files are replicated at more than one site, it occupies more storage space. This requires more additional maintenance. This becomes an overhead in storing multiple files.
- Lower write performance: Performance of write operations can be considerably lower in applications requiring high updates in replicated environment, because the transaction may need to update multiple copies.

The main aim of using replication is to reduce access latency and bandwidth consumption. The other advantages of replication are that it helps in load balancing and improves reliability by creating multiple copies of the same data. These can generally be classified as static and dynamic. In static replication, a replica exists until it is deleted by users or its duration is expired. Static replication is used to copy data to other datacenter where it is most popularly requested. The drawback of static replication is evident when client access patterns change greatly in the data. The drawback with static replication is that it cannot adapt to changes in user behavior.

4. CLOUD BROKER–I NTRODUCTION

A Cloud Broker is a third-party individual or business that acts as an intermediary between the Cloud users and Cloud service provider. In general, a broker is someone who acts as an intermediary between two or more parties during negotiations [11]. The broker's primary role may simply be to save the user's time by researching services from different providers and helps users to get the best provider for their requirement. Once the broker completed his search, he presents the customer with a short list of recommended cloud providers and the customer contacts the providers of choice to arrange service. A cloud broker may also be granted the rights to negotiate contracts with cloud providers on behalf of the users. In large scale organizations, the cloud brokers are given rights manage the services, monitoring the users and so on.

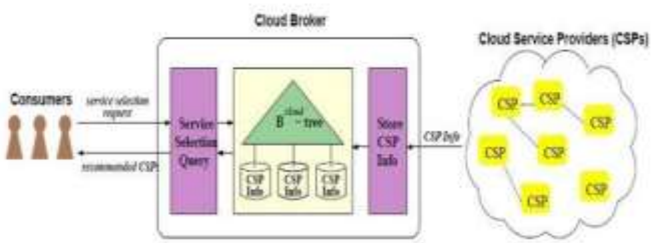

Fig 3: - Cloud Broker Model

4.1. Types of Cloud Brokers

Cloud Brokerage (CB) is a brokering service in which any number of Cloud services can be delivered to a user. Services can range from the one time registration in SaaS, to integrating the data between all other layers such as PaaS and IaaS. i.e., an ERP across SaaS layer can be added to a management layer across IaaS layers.

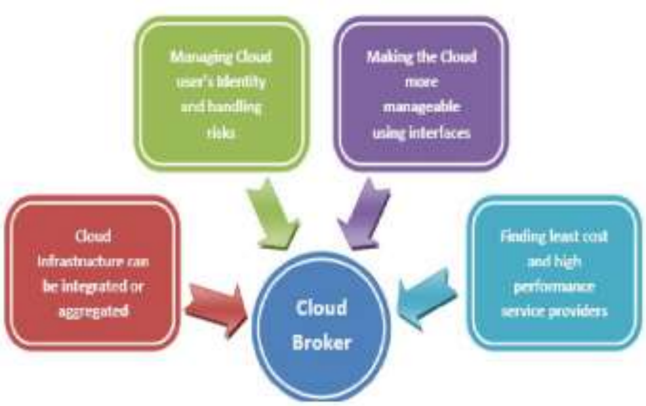

Fig 4: - Types of Cloud Brokers

There are three types of brokering services which fall into the patterns according to Gartner.

- **Cloud Aggregator:** Some Cloud brokers are involved in collecting the detailed information about the services of Cloud providers and adding capabilities by improving some aspects of those services. They are also involved in adding and managing the hosting services.
- **Cloud customizer:** This type of Cloud broker mainly focus on customizing multiple cloud services from different cloud providers and integrating together for an organization.
- **Brokers that arbitrage cloud services:** These brokers helps the Cloud users to choose several cloud providers for their requirements that are dependent on attributes such as costs or performance

5. QOS BROKERING ALGORITHM – ANALYSIS

Cloud Computing is nowadays one of the most popular computational paradigms. It has been adopted by many companies and considered by many more others for the unquestionable benefits offered, such as potential cost reductions offered by the pay-per-use model, flexibility and scalability, fault-tolerance and increased availability for the geographic distribution of resources [12]. However, most Cloud providers force their customers to use proprietary interfaces, virtualization technologies, and communication protocols and so on. This may lead to a high degree of vendor lock-in that occurs when a customer is tightly-coupled to a single

5.1. QOS Broker Resource Allocation Algorithm

A QoS broker receives clients' functional & QoS requirement requests and identifies qualified services for them. Its main components include QoS information manager, QoS negotiation manager and QoS analyzer.

QoS broker acts as the front-end of a server. In the specialized architecture, we combine QoS information module in server with QoS information manager module in broker. In addition, the QoS admission & enforcement module is combined with the QoS negotiation manager [13] [14]. These two new modules (QoS Information Manager and QoS Admission & Enforcement) all reside in the broker, i.e. the broker controls admission for the server and determines how much resource should be allocated to each client.

We study two resource allocation algorithms, HQ and RQ, which are used for legacy and QoS servers respectively. We improve the algorithms by adjusting the allocation policy according to the feedback of historical statistical information. The feedback is very useful since they reflect the client behavior and the performance of the algorithms [15]. Based on the information, we can have more efficient policies as time goes by. Many existing Web service providers utilize legacy servers that have no control on service quality. In that case, clients must rely on the broker to ensure the desired QoS level. For such servers, a broker uses a homogeneous resource allocation for QoS admission and assignment.

5.2. Genetic Algorithms

Genetic Algorithm (GA) is a well-known heuristic approach that permits to iteratively find near-optimal solutions in large search spaces. Our work leverages the GA approach because it is flexible enough to support multiple constraints and the injection of additional constraints with minimal interventions on the algorithm. Clearly, this is crucial for software reuse in the context of Cloud Computing, where QoS models are continuously enriched as providers support QoS guarantees previously not addressed, such as soft real-time guarantees for virtualized services

5.3. The RPBP algorithm

The RPBP algorithm selects cloud physical resources to host a VM using simple profiling algorithms. The profiling is based on either the CPU capacity or the historical network delay. The profiles are used to rank the resources then the highest rank resources are selected subject to the budget constraints.

5.4. Hybrid Algorithm

A hybrid cloud scheduler decides (based on the budget and deadline) whether a task can be scheduled on the private cloud or if it should be transferred to the public cloud [16]. The tasks are stored in a queue based on their deadlines. A queue scanning algorithm is used to detect tasks that cannot be satisfied within the deadline (based on the elapsed and remaining time of the tasks in the queue). These tasks will then be executed on the public cloud.

5.5. The QBrokerage algorithm

The QBrokerage algorithm is proposed based on a genetic algorithm to select cloud providers. First, the cloud application is modeled as a directed graph in which the vertices represent virtual machines and edges represent the communication paths between virtual machines. Second, the graph is passed to a mapping component which searches for services that satisfy the requirements of the VM.

The Performance Analysis of the various brokering algorithm is given below.

Brokering Algorithm	Time	Security	Complexity	Availability	Customer Service
QOS Broker Resource Allocation Algorithm	Y	X	Y	Y	Y
Genetic Algorithm	Y	Y	Y	Y	Y
The RPBP Algorithm	Y	Y	X	Y	Y
Hybrid Algorithm	Y	Y	Y	Y	Y
The QBrokerage algorithm	X	Y	Y	Y	Y

Table 1: - Performance Comparison

6. CONCLUSION

Here, addressed two issues related to placing replicas of the objects in multi-cloud environment. In order to tackle these issues, we propose efficient algorithms. The first algorithm has been designed to minimize the replication cost and the expected availability of objects as the user's QoS is met. The second one is proposed to maximize the expected availability of objects under a given budget with the assumption that the objects are split to chunks.

As our future work, first propose an algorithm to find the minimum replication cost with a given expected availability for striped objects. Second, since in this work we have conducted trade-off between the storage cost and the availability, The main aim to consider the data transfer in/out cost, the queries and the processing cost when we select the cloud provider. Third, the cost of migration from one cloud provider to another due to urgent needs should also be taken into consideration in selection of cloud storages.

7. REFERENCES

[1] Slimani, S., Hamrouni, T., & Ben Charrada, F. (2021). Service-oriented replication strategies for improving quality-of-service in cloud computing: a survey. Cluster Computing, 24, 361-392.

[2] Malik, S. U. R., Khan, S. U., Ewen, S. J., Tziritas, N., Kolodziej, J., Zomaya, A. Y., ... & Li, H. (2016). Performance analysis of data intensive cloud systems based on data management and replication: a survey. Distributed and Parallel Databases, 34, 179-215.

[3] Hamrouni, T., Mokadem, R., & Khelifa, A. (2023). Review on data replication strategies in single vs. interconnected cloud systems: Focus on data correlation-aware strategies. Concurrency and Computation: Practice and Experience, 35(22), e7758.

[4] da Silva Veith, A. (2019). Quality of service aware mechanisms for (re) configuring data stream processing applications on highly distributed infrastructure (Doctoral dissertation, Université de Lyon).

[5] Sabaghian, K., Khamforoosh, K., & Ghaderzadeh, A. (2023). Data Replication and Placement Strategies in Distributed Systems: A State of the Art Survey. Wireless Personal Communications, 129(4), 2419-2453.

[6] Mansouri, Y. (2017). Brokering algorithms for data replication and migration across cloud-based data stores (Doctoral dissertation, University of Melbourne, Parkville, Victoria, Australia).

[7] Jawarneh, A., & Hasan, I. M. (2020). Quality of service aware data stream processing for highly dynamic and scalable applications..

[8] Jawarneh, A., & Hasan, I. M. (2020). Quality of service aware data stream processing for highly dynamic and scalable applications.

[9] Mansouri, N., & Javidi, M. M. (2020). A review of data replication based on meta-heuristics approach in cloud computing and data grid. Soft computing, 24(19), 14503-14530.

[10] Mansouri, N., & Javidi, M. M. (2018). A new prefetching-aware data replication to decrease access latency in cloud environment. Journal of Systems and Software, 144, 197-215.

[11] Asghari, P., Rahmani, A. M., & Javadi, H. H. S. (2022). Privacy-aware cloud service composition based on QoS optimization in Internet of Things. Journal of Ambient Intelligence and Humanized Computing, 13(11), 5295-5320.

[12] Dabas, C., & Aggarwal, J. (2019). An intensive review of data replication algorithms for cloud systems. In Emerging Research in Computing, Information, Communication and Applications: ERCICA 2018, Volume 1 (pp. 25-39). Springer Singapore.

[13] Nannai John, S., & Mirnalinee, T. T. (2020). A novel dynamic data replication strategy to improve access efficiency of cloud storage. Information Systems and e-Business Management, 18(3), 405-426.

[14] Salem, R., Salam, M. A., Abdelkader, H., & Mohamed, A. A. (2019). An artificial bee colony algorithm for data replication optimization in cloud environments. IEEE Access, 8, 51841-51852.

[15] Veith, A. D. S. (2019). Quality of Service Aware Mechanisms for (Re) Configuring Data Stream Processing Applications on Highly Distributed Infrastructure (Doctoral dissertation, Université Rennes 1).

[16] Venkataramanachary, V., Reveron, E., & Shi, W. (2020, January). Storage and rack sensitive replica placement algorithm for distributed platform with data as files. In 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS) (pp. 535-538). IEEE.