IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

A Survey of Security, Privacy Preservation, and Service Quality in Cloud Computing Environments

Dr. Puneet Garg
Associate Professor, Department of CSE-AI
KIET Group of Institutions, Delhi NCR, Ghaziabad
puneet.garg@kiet.edu

Abstract: *Cloud computing has transformed modern information technology by providing scalable, on-demand, and cost-effective computing resources through the Internet. Despite its widespread adoption across diverse domains, concerns related to security, privacy preservation, and Quality of Service (QoS) continue to hinder its full potential. This survey presents a comprehensive review of cloud computing environments by examining deployment and service models, enabling technologies, and the major challenges associated with secure and efficient cloud service delivery. The paper discusses fundamental security requirements, including authentication, authorization, confidentiality, integrity, non-repudiation, and availability, as well as prominent cloud security models. It further reviews privacy-preserving mechanisms that support secure data access, confidentiality, owner-controlled data sharing, and integrity verification in untrusted cloud environments. In addition, the study explores QoS by analyzing its key parameters, management techniques, and service quality attributes that influence cloud performance and Service Level Agreement (SLA) compliance. A comparative review of recent literature highlights current research trends, strengths, limitations, and emerging solutions involving artificial intelligence and machine learning for cloud optimization. The survey identifies existing research gaps and provides valuable insights for designing secure, privacy-aware, and high-performance cloud computing systems capable of addressing future technological and operational demands.*

Keywords: Cloud Computing, Security, Privacy Preservation, Quality of Service (QoS), Data Security, Encryption, Resource Management, Service Level Agreement (SLA), Cloud Services

I. INTRODUCTION

In a cloud computing environment there are two components playing an important role these are cloud providers and cloud users. From the perspective of a cloud provider, the providers have a large number of computing resources in their large data centers and they rent out these resources to the users on a pay-per-use basis to maximize the revenue by attaining [1]. Resources are also in demand for the cloud users and applications with dynamic nature which are predicted by them. For the cloud users, who have applications with fluctuating loads lease the resources from the cloud providers and run their applications within minimum expenses [2]. Every cloud user wants a number of resources for a particular task or cloudlet that can maximize the performance and have to be finished on time[3]. Security is one of the biggest obstacles that hamper the widespread adoption of cloud computing. Several business and research organization are reluctant in completely trust the cloud computing to shift digital assets to the third-party service providers The conventional IT infrastructure keeps the digital assets in the administrative domain of the organizations [4]. The security measures taken by the cloud service providers (CSP) are generally transparent to the organizations. The

ISSN: 2581-9429
Impact Factor: 8.2

presence of large numbers of users that are not related to the organizations, aggravate the concerns further the users might be trusted by the CSP but they may not be of trust to each other [5][6].

Privacy deals with accessibility and availability of sensitive information to the intended recipients. Outsourced data is accessed and modified only by the users who have appropriate access privileges [7][8]. Consider an organization have prepared and outsource their data in cloud. The outsourced data is provided to the local administrator for storage in the cloud. It is recommended to ensure that the administrator cannot view or modify the data contents. After outsourcing the file, no one, including the service provider, can view or modify its contents. If the service provider or local administrator is trying to read the file contents, it must not be done. Privacy preservation concerns the security of outsourced data. providing dedicated cloud services that ensure users' dynamic quality of service requirements and avoid Service-Level Agreement (SLA) violations, along with a cost-effective way [9], is still one of the biggest challenges for the cloud providers' side. SLA specifies quality of service targets and economic penalties associated to SLA violations. Common quality of service include availability, response time, security, latency, reliability, etc [10][11].

A. Structure of Paper

This paper is organized as follows. Section II presents an overview of cloud computing. Section III discusses security mechanisms, Section IV reviews privacy preservation techniques, and Section V explores service quality in cloud computing. Section VI provides a comparative analysis of recent studies, while Section VII concludes the paper and outlines future research directions.

II. OVERVIEW OF CLOUD COMPUTING

The cloud computing environment and developing standards and guidelines which shows a valuable contribution that offers a better understanding of cloud services and computing technologies [12].

A. Cloud Deployment Models

The cloud computing model has three deployment models that can be particularly used to represent the cloud service models . The four deployment models: public cloud, private cloud, community cloud, and the hybrid cloud. The deployment models represent the way that the computing infrastructure delivers the cloud services can be employed.

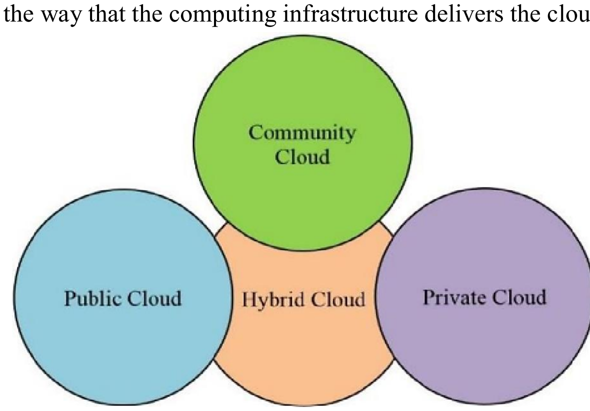

Fig. 1.Cloud deployment models

1) **Public Cloud**: A public cloud represents the cloud hosting and owned by the service provider whereby the client and resource provider have service level agreement [13]. Microsoft, Google, Amazon, VMware, IBM, Sun and Rackspace are some examples of cloud service provider. The plateform is designed in the form of generalized computing that holds the generic type of customer demand. The resources are made available to the public and easily accessible. Multiple entities are involved in operating public cloud and resource are public for the customers which makes them difficult to protect from malicious attacks. It contains some concerns over privacy, data access and security for

DOI: 10.48175/IJARSCT-37335

customers because it is outside the firewall. It is less secure than the other deployment models and suited for a small and medium business that may not have to configure servers and purchase capital resources.

2) **Private Cloud:** The cloud infrastructure is managed and maintained by the single organization that compromises multiple customers. If any organization set up their own private cloud and recently create their own servers having physical hardware servers that put virtualization layer top on them then they would make resources available only internally. So, their application can deploy to their own physical control server, they don't need to go Microsoft or Amazon servers. It can ensure the physical security and more secure as compared to the public cloud because of its specific internal exposure. However, the cost is significantly higher because expertise and training are needed for the server administrator, virtualization specialist, and network specialist. In private cloud, it is easier to address the relationship between the service provider and customer because the infrastructure operated and owned by the same organization [14]. It employs the capabilities of cloud management software to ensure reliable delivery service and integrity of the external resources.

3) **Hybrid Cloud**: Hybrid cloud is referred as the combination of two or more cloud deployment models that can be either public, private or community clouds which remains the unique entities but are bound together [14]. The importance of hybrid cloud usually offers extra resources when the high demand from the customer and for instance it is enabled to migrate some computation jobs from private to public cloud. It is well organized and allow different entities to access data over the internet because it offers more secure control of the applications and data. It provides a benefit over different deployment models and can be internally and externally hosted. Hybrid cloud gets more popularity and became a dominant model. The main reason is that it has the ability to take advantage of cost-saving, scalability in elasticity that public cloud may provide, allow control flexibility when it needed [15].

4) **Community Cloud**: Community cloud is referred as the organizations shared its cloud infrastructure among the customers having similar interest or concerns like a policy, the security requirements, mission and compliance consideration [16]. Several organizations or a third party operate, control, share and handle the resources of the community cloud [17]. In case of the third party like Siemens have IT services and solutions that set up a media cloud for the media industry. It tends to be rarer and more specialized. The cloud infrastructure of community cloud is shared and owned by different organizations such as research groups, together with work of companies and government organizations [18].

B. Cloud Computing Service Models

The three cloud service models or delivery models are available for the customer: The models are built in order to satisfy the requirements of the client. The models are Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). There are different levels of security required for these service models in the cloud environment.

Infrastructure as a service (IaaS)

The service offers a way of acquiring computing capabilities over the web. The client can utilize resources such as processing power, storage facilities, VPN, and networks. The client is charged by a "pay-as-you-go" model, where the customer is billed depending upon the use of the processing power and storage space. The IaaS services are utilized on platforms such as Google Cloud Platform and Amazon EC2

Platform as a Service (PaaS)

The Cloud Provider gives the ability to the client in order to deploy customer created applications using tools and programming languages. The provider does not control the underlying architecture including operating system, storage and servers, etc. The provider presents a cloud-based environment in which the user can build applications without the need of working with Integrated Development Environments (IDEs). For example: Google allows the clients to make their own applications using the Google App Engine. Other examples include Microsoft Azure and Apache Startos.

Software as a service (SaaS)

The cloud provider leases the software or applications which are owned by them to the client. For example: Customer Relation Manager (CRM) is provided by salesforce.com on a cloud infrastructure to its client. The company charges the client for its services, but the software is owned by the salesforce company (software provider). This helps the organizations since they do not have to maintain the software or secure the network, and other operational issues which keep the applications running.)[19]

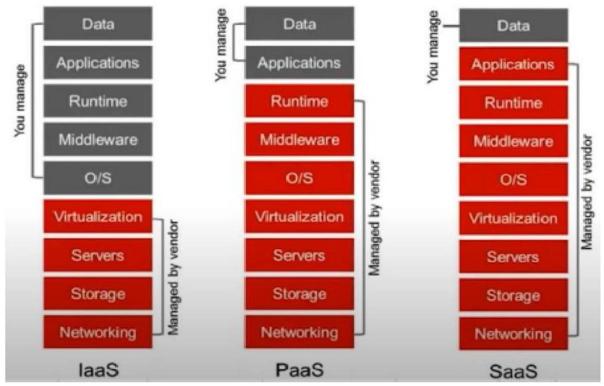

Fig. 2.Cloud computing service models

C. Enabling Technologies of Cloud Computing

Cloud Computing distinguishes itself from other computing paradigms like grid computing, global computing, and internet computing in various aspects of on demand service provision, user centric interfaces, guaranteed QoS (Quality of Service), and autonomous system [20] etc. A few state-of-the-art techniques that contribute to cloud computing are:

- **Virtualization**: It has been the underlying concept towards such a huge rise of cloud computing in the modern era. The term refers to providing an environment that is able to render all the services, supported by a hardware that can be observed on a personal computer, to the end users. The three existing forms of virtualization categorized as: Server virtualization, Storage virtualization and Network virtualization, have inexorably led to the evolution of Cloud computing [21] .
- **Web Service and SOA**: Web services provided services over the web using technologies like XML, Web Services Description Language (WSDL), Simple Object Access Protocol (SOAP), and Universal Description, Discovery, and Integration (UDDI). The service organisation inside a cloud is managed in the form of Service Oriented Architecture (SOA) and hence it can define SOA as something that makes use of multiple services to perform a specific task [22] .
- **Application Programming Interface (API)**: Without APIs, it is hard to imagine the existence of cloud computing. The whole bunch of cloud services depend on APIs and allow deployment and configuration through them. Based on the API category used, viz. control, data and application, different functions of APIs are invoked and services are rendered to the users accordingly.
- **Web 2.0 /Mash-up**: Web 2.0 has been defined as a technology that enables us to create web pages and allows the users to interact and collaborate as creators of user generated content in a virtual community [23]. It enables the usage of World Wide Web technology towards a more creative and a collaborative platform [24]. Mash-up is a web application that combines data from more than one source into a single integrated storage tool [25].

III. SECURITY IN CLOUD COMPUTING

The cloud computing security encompasses the policies, requirements, and architectural models that protect cloud resources, applications, and data from unauthorized access and cyber threats. It ensures confidentiality, integrity,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

availability, authentication, and accountability while supporting secure service delivery through robust security frameworks, access control mechanisms, and standardized cloud security models.

A. Information Security Requirements

Cloud computing security should also be guided in this regard in order to become an effective and secure technology solution. illustrating the information security requirements coupled [26]. In the different cloud delivery models and deployment models are matched up against the information security requirements. Cloud computing. should be viewed in context as a guideline in assessing the security level. Each of the security requirements will be highlighted below in context of Cloud computing.

Identification & authentication:

In Cloud computing, depending on the type of cloud as well as the delivery model, specified users must firstly be established and supplementary access priorities and permissions may be granted accordingly. This process is targeting at verifying and validating individual cloud users by employing usernames and passwords protections to their cloud profiles.

Authorisation

It is an important information security requirement in Cloud computing to ensure referential integrity is maintained. It follows on in exerting control and privileges over process flows within Cloud computing. Authorisation is maintained by the system administrator in a private cloud.

Confidentiality

In Cloud computing, confidentiality plays a major part especially in maintaining control over organisations' data situated across multiple distributed databases. It is a must when employing a public cloud due to public clouds accessibility nature. Asserting confidentiality of users' profiles and protecting their data, that is virtually accessed, allows for information security protocols to be enforced at various different layers of cloud applications.

Integrity

The integrity requirement lies in applying the due diligence within the cloud domain mainly when accessing data. Therefore ACID (atomicity, consistency, isolation and durability) properties of the cloud's data should without a doubt be robustly imposed across all Cloud computing deliver models

Non-repudiation:

Non-repudiation in Cloud computing can be obtained by applying the traditional e-commerce security protocols and token provisioning to data transmission within cloud applications such as digital signatures, timestamps and confirmation receipts services (digital receipting of messages confirming data sent/received).

Availability

Availability is one of the most critical information security requirements in Cloud computing because it is a key decision factor when deciding among private, public or hybrid cloud vendors as well as in the delivery models. The service level agreement is the most important document which highlights the trepidation of availability in cloud services and resources between the cloud provider and client. Therefore, by exploring the information security requirements at each of the various cloud deployment and delivery models set out by the ISO, vendors and organizations can become confident in promoting a highly protected safe and sound cloud framework[27] .

B. The Security Models of Cloud Computing

The Cloud Multiple-Tenancy Model of NIST

Multiple-tenancy is an important function characteristic of cloud computing that allows multiple applications of cloud service providers currently running in a physical server to offer cloud service for customers. This physical server partitions and processes different customer demands with virtualization. Virtualization possesses good capability of sharing and isolation, and is a right core technology of cloud computing. By running multiple virtual machines (VMs) in a physical machine, virtualization enables to share computing resource such as processor, memory, storage, and I/O among different customers' applications, and improves the utilization of cloud resources.

Jerico Formu's Cloud Cube Model

Jerico formu's cloud cube model is a security attribute information implied in the service and deployment models of cloud computing and the location [28], manager and owner of computing resources. It's components are:

- Internal/External: A model parameter to define the physical location of data storage. If the physical location of data storage is inside of the data owner's boundary, then the model parameter value is internal.
- Proprietary/Open: Proprietary means that a cloud service provider holds the ownership of facilities providing cloud services, hence the operation of proprietary and customers can not transfer their applications from one to another cloud service provider without great effort or investment.

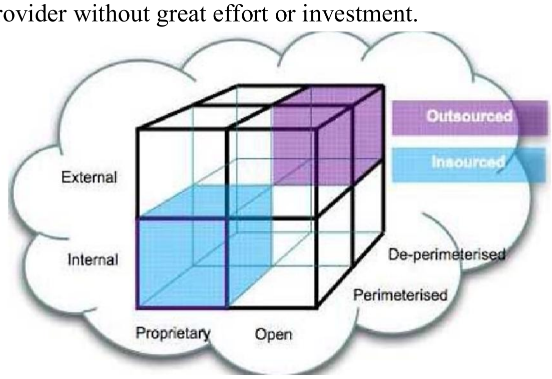

Fig. 3. Cloud cube model of Jericho forum

- Perimeterised/De-perimeterised: Perimeterised means that a customer's application operates within traditional IT security boundary signaled by firewall that blocks the incorporation of different security zones. In fact, customers running some applications inside of security zone can extend/shrink their application perimeter to/back from external cloud environment by VPN.De-perimeterised means that the fadeway of traditional IT security boundary and the exposure of a customer's application operation.
- Insourced/Outsourced: A model parameter to define the 4th dimension that has two states in each of the eight cloud forms: Per (IP, IO, EP, EO) and D-p (IP, IO, EP, EO). Insourced means that cloud service is presented by an organization's own employees, and Outsourced means that cloud service is presented by a third party.

IV. PRIVACY PRESERVATION IN CLOUD COMPUTING

The cloud computing privacy focuses on protecting sensitive user data throughout its storage, processing, sharing, and access. It employs secure data access mechanisms, confidentiality-preserving encryption, owner-controlled data sharing, and integrity verification techniques to safeguard privacy, prevent unauthorized disclosure, and maintain trust in cloud-based services and applications.

Privacy-Preserving Cloud Data Access.

User privacy in cloud services includes identity privacy, query privacy, and access pattern privacy. Different privacy issues can be addressed with different protection techniques and need some specific considerations. To achieve access pattern privacy, three typical protection approaches are proposed: Private Information Retrieval (PIR)and dynamically allocated data structure. Query privacy can be effectively implemented by using secure indexes virtual/dummy keywords or random trapdoors When data is shared within a group, ring or group signature techniques are used to keep user identity information secret.

Confidentiality-Assured Cloud Data Service.

In public cloud environments, it is necessary to keep data secret while providing normal data utilization services. Encryption is a basic mechanism to enable data confidentiality. However, traditional encryption hampers the data utilization services. Intuitively, all the outsourced data may be downloaded and decrypted locally to perform data search or data computation. However, this naive solution is obviously impractical for the reason of prohibitive communication overheads. To achieve confidentiality-assured and effective data services in the cloud, new cryptographic primitives and data encryption proposals have been presented.

Owner-Controlled Cloud Data Sharing.

In untrusted cloud environments, a challenging problem is to enable fine-grained enforcements of data access and achieve secure data sharing among large-scale users. Obviously, since an owner does not trust the cloud, traditional access control mechanisms, normally depending on a trusted server, are not suitable for cloud data sharing. To address this challenge and enforce owner-controlled access control, data should be encrypted by the owner before outsourcing it to the cloud, and then the owner can perform fine-grained access control over the encrypted data by securely distributing keys.

Integrity-Guaranteed Cloud Data Storage.

Since outsourcing data to the cloud implies that owners no longer locally possess their data, owners are concerned whether the outsourced data is correctly stored in the cloud. Traditional approaches of data correctness guarantees cannot be adopted directly, because a local data copy is generally required to verify the data integrity. Furthermore, the limited computing resources of the owners and the vast amount of cloud data make the task of checking data intactness expensive and even unbearable. Aiming at efficiently verifying data integrity in a cloud, blockless verification and probabilistic check have been adopted. Blockless verification allows a user to verify data authenticity without accessing the actual data, while probabilistic check enables a user to verify data integrity by launching a random challenge-response protocol [29].

V. SERVICE QUALITY IN CLOUD COMPUTING

Cloud Computing aims to distribute a portfolio of service in the virtual environment that consumers can access them from anywhere in the world at competitive costs depending on their QoS requirements [30]. QoS is necessary for Cloud Costumers, who ask providers to deliver for them the advertised quality characteristics, and for Cloud providers, who look to find the right tradeoffs between QoS levels and operational cost. Nonetheless, consumers can select a Cloud service appropriate for them from major service providers that can offer services with QoS [31].

A. Quality of Service (QoS) Parameters

The QoS parameters to determine properties of the service. They can be divided into qualitative and quantitative attributes. To evaluate the Cloud services performance in terms of a large set of QoS parameters such as:

1) Throughput (TP) is the maximum requests that can be handled at a given unit in time, expressed in requests per minute.
2) Latency (L) is the time it takes for data service packets to be stored or retrieved expressed in milliseconds.
3) Jitter (J) is defined as a variation in the delay of received packets expressed in milliseconds.
4) Response time (RT) is the time taken to deal with a request by the processor and submit the response, expressed in milliseconds.
5) Availability (A) is the probability that a web service is available, expressed in Available time per total time.
6) Reliability (RL) is the probability that a task is satisfactorily fulfilled.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

B. Quality of Service (QoS) Management Techniques in Cloud Computing

Quality of Service (QoS) management focuses on ensuring reliable, efficient, and scalable cloud service delivery by maintaining optimal performance, maximizing resource utilization, meeting Service Level Agreements (SLAs), minimizing service disruptions, and enhancing the overall user experience in dynamic cloud environments.

1) Resources management

The resource allocation problem occurs when the provider wants to execute a suitable number of VMs on appropriate physical servers [32][33]. The mechanisms of resource allocation is to assign resource shares trying to minimize management costs and grantees the QoS.

2) Admission control mechanism

Admission control is an overload protection mechanism for two type infrastructures. First, the infrastructure-provider that the QoS admission control mechanism rejects requests under peak workload conditions to prevent QoS degradation. The admission controller rejects all new sessions favoring the service of requests from already admitted sessions, if such value grows above a certain threshold, this decision is taken by predicting the value of a specific QoS metric. Second, about the infrastructure-user the admission control mechanism is used as an extreme overload mechanism, helpful when additional resources are obtained with some significant delay.

3) Scheduling mechanisms

Scheduling in Cloud Computing is the selection of best appropriate resources for task execution. A task scheduler in Cloud Computing has to satisfy Cloud users with the QoS requirements and improve profits of Cloud providers. The scheduling involves the selection of the services and the appropriate time when task started.

4) Performance models

Performance modeling approaches are used to evaluate and predict Quality of Service (QoS) in cloud computing environments. They support resource provisioning, workload balancing, and capacity planning [34]. Common approaches include queuing systems, queuing networks, and layered queuing networks, which model resource behavior, application interactions, and multi-layer dependencies to improve performance prediction and QoS management [35].

5) Monitoring mechanisms

Cloud monitoring is the dynamic tracking technique of the QoS parameters (running or hosted on the VM) associated to virtual Cloud services. This mechanism can help all professional users (developer, administrator): to keep the Cloud applications and services running at the highest efficiency, to keep track of the SLA violations, to detect the performance of certain QoS parameters, to account the percentage of leave and join operation of services and detect the phenomenon caused these operations.

C. Quality of Service (QoS) Attributes

Quality of Service (QoS) attributes define the key characteristics used to evaluate cloud service performance, dependability, and usability, enabling providers to maintain service quality while ensuring reliable, scalable, and efficient user experiences.

- **Availability:** This attribute assesses the ratio of the total time to the time which a web service is capable of being functional. Typically, availability objectives are determined either in decimal fractions or percentage. The principal for defining this attribute is that web service to be available is a requirement for the successful execution of the web service. Unless high availability of services is not assured, service consumers cannot use the services on demand.
- **Reliability**: This attribute assesses the ability of the web service to keep functioning with particular level of performance over time. Various facets of reliability are important within a cloud computing, especially the reliability of the messages that are transferred between the service consumers and the services. Therefore, the reliability of web service has an effect on service users themselves. The execution of the services is done on service provider's side.

- **Scalability:** This attribute assess the ability to either manage increasing amounts of resources, used for Web services or to be readily elaborated. This attributed is mainly defined for the reason is that cloud services are provide their services dynamically scalable manner and frequently virtualized resources are rendered as web services via internet. If the service providers render their application which can scale resources flexibly, they can get more benefit in terms
- **Efficiency:** Efficiency measures the resource utilized for web services when rendering the required functionality and the level of performance, under stayed condition, which means it assess how good the web services use the resources. Cloud computing is a typical business model depends on the amount of utilized resource or time-use.
- **Reusability:** This measures the extent to which a cloud service can satisfy the requirements of multiple consumers across different applications. Highly reusable services increase adoption, improve return on investment (ROI) for providers, and reduce development effort for consumers. It also considers service understandability through clear documentation and service visibility, enabling users to efficiently discover, evaluate, and integrate available services.
- **Adaptability and Composability**: Adaptability and composability measure a service's ability to integrate, adapt, and function effectively across diverse cloud environments and service-based applications. These attributes enhance interoperability, enabling services to be easily combined, reused, and modified to meet changing application requirements while maintaining seamless communication and functionality.

VI. LITERATURE REVIEW

This section reviews recent studies on cloud computing security, privacy preservation, and Quality of Service (QoS). The recent studies are given:

Nallabollu et al., (2026) perform investigation on parameter-efficient fine-tuning (PEFT) techniques, including adapters, prefix-tuning, and low-rank adaptation (LoRA) for adapting LLMs to cloud-based data pipelines under constrained conditions. The results show that PEFT approaches can reduce trainable parameters by more than 90% while retaining up to 98% of the task accuracy, thereby enabling deployments to scale. They further propose a lightweight hybrid resource-aware strategy that dynamically allocates fine-tuning methods based on cloud constraints. These findings demonstrate that PEFT not only democratizes LLM adaptation but it also offers a viable path toward cost-efficient, cloud-based AI services in the resource limited environments [36].

Suthari et al., (2025) explores machine learning-based methods for demand forecasting and adaptive scheduling to improve cost efficiency. Through experiments on synthetic and real-world datasets, the proposed framework demonstrates significant savings in cloud ownership costs, achieving up to 24.6% reduction, while maintaining system reliability and meeting performance and SLA requirements. The findings provide valuable insights for practitioners seeking to balance cost-efficiency and performance in cloud infrastructures. This work contributes to advancing cost optimization strategies, particularly for organizations with dynamic workloads [37].

Vij and Mishra (2025) suggests a clever paradigm for resource optimization that combines carbon footprint reduction, workload prediction, and hyper-scheduling across cloud and edge nodes. The framework uses cutting-edge machine learning methods, such as reinforcement learning and deep neural networks, to dynamically modify system settings in real time. Comparing the suggested strategy to traditional resource allocation techniques, simulation tests show that it greatly improves operational responsiveness, energy stability, and service efficiency. The findings demonstrate the technical potential of AI-driven optimization in creating computing ecosystems that are self-adaptive and sustainable [38].

Hassan et al., (2024) highlights the role of artificial intelligence in the cloud computing environment. Different machine learning algorithms that help in cloud computing tasks have also been explored. Major machine learning algorithms for cloud computing are K-means, SVM, K-NN, Naïve Bayes, SVD, and neural networks. The major research trends have also been explored in which cloud security is the most researched theme of cloud computing. The applications of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

artificial intelligence in the cloud environment make the cloud environment more cost-effective, secure, and efficient [39].

S and Ponmagal (2023) proposes Transformer-enhanced Distributed DRL scheduling technique, called TF-DDRL, to adaptively schedule heterogeneous IoT applications. This technique follows the Actor-Critic architecture, scales efficiently to multiple distributed servers, and employs an off-policy correction method to stabilize the training process. In addition, Prioritized Experience Replay (PER) and Transformer techniques are introduced to reduce exploration costs and capture long-term dependencies for faster convergence. Extensive results of practical experiments show that TF-DDRL, compared to its counterparts, significantly reduces response time, energy consumption, monetary cost, and weighted cost by up to 60%, 51%, 56%, and 58%, respectively [40]

Karabetian et., al (2022) present a solution on dynamic resource allocation for efficient cloud scalability. This is made possible by using machine learning algorithms as well as user feedback, in order to generate an adequate resource forecasting model. The efficiency of the tool is evaluated by repeatedly executing extensive analysis of various datasets provided by the end users, exploiting the cloud computing paradigm for their analytic purposes. The solution is able to learn and enhance its knowledge graph considering user feedback, as well as previously executed processes in cloud environment. To this extent, the forecasting model will attempt to estimate optimal resource allocation for each user scenario [41].

Table 1: Comparative Analysis of Recent Study on Cloud Computing Environments

Reference	Approach	Findings	Advantages	Challenges	Recommendations
Nallabollu et al. (2026)	Parameter-Efficient Fine-Tuning (LoRA, Adapters, Prefix-Tuning)	Reduced trainable parameters by 90% while retaining 98% accuracy	Enables scalable and cost-efficient cloud AI services	Performance depends on resource-constrained environments	Investigate PEFT for heterogeneous and multi-cloud platforms
Suthari et al. (2025)	ML-based resource allocation using overcommitment, workload consolidation, auto-scaling, spot and reserved instances	Reduced cloud ownership cost by 24.6% while satisfying SLA requirements	Improves cost efficiency and resource utilization	Dynamic workload prediction remains complex	Extend the framework to serverless and edge computing environments
Vij and Mishra (2025)	AI-based hyper-scheduling with reinforcement learning and deep neural networks	Improved service efficiency, energy stability, and reduced carbon footprint	Sustainable and adaptive cloud resource management	Requires high computational resources and real-world validation	Develop lightweight AI models for large-scale cloud systems
Hassan et al. (2024)	AI and Machine Learning techniques for cloud optimization	AI improves security, resource allocation, and cost optimization	Enhances cloud intelligence and automation	AI model complexity and computational overhead	Investigate explainable AI and autonomous cloud management
S and Ponmagal (2023)	Transformer-enhanced Distributed Deep Reinforcement	Reduced response time, energy	Efficient scheduling and better QoS	Training complexity and scalability issues	Evaluate under large-scale real-world cloud workloads

	Learning (TF-DDRL) scheduling	consumption, and execution cost			
Karabetian et al. (2022)	Machine learning-based dynamic resource forecasting	Improved cloud scalability and resource allocation accuracy	Adaptive resource management using user feedback	Requires diverse datasets for accurate prediction	Develop more robust forecasting models for heterogeneous clouds

VII. CONCLUSION AND FUTURE WORK

Cloud computing has become a fundamental technology for delivering scalable, flexible, and cost-efficient computing services to individuals and organizations. However, maintaining a balance between security, privacy preservation, and Quality of Service (QoS) remains a significant challenge due to the dynamic and distributed nature of cloud environments. This survey reviewed the core concepts of cloud computing, including deployment and service models, enabling technologies, security requirements, privacy-preserving mechanisms, QoS management techniques, and QoS attributes. It also summarized recent research advances, highlighting the growing adoption of artificial intelligence and machine learning for intelligent resource allocation, cloud security enhancement, cost optimization, and performance improvement. The comparative analysis revealed that although considerable progress has been achieved, challenges such as evolving cyber threats, privacy risks, resource management complexity, interoperability, and SLA compliance continue to require further attention. Future research should focus on developing lightweight and explainable AI-driven security frameworks, privacy-preserving data analytics, adaptive trust management, secure multi-cloud and edge-cloud integration, and energy-efficient resource management strategies. Furthermore, integrating blockchain, zero-trust architectures, federated learning, and privacy-enhancing technologies with intelligent QoS optimization can significantly improve cloud reliability, security, and user trust. These advancements will contribute to building resilient, sustainable, and secure cloud computing environments capable of meeting the increasing demands of next-generation digital applications.

REFERENCES

[1] A. Joon, B. K. R. Janumpally, A. Gogineni, and P. Chatterjee, “Efficient Large-Scale Intrusion Identification and Prevention in Distributed Cloud Networks Using Artificial Intelligence,” in *2025 5th International Conference on Intelligent Technologies (CONIT)*, IEEE, Jun. 2025, pp. 1–8. doi: 10.1109/CONIT65521.2025.11167760.

[2] V. K. Sharma, “Cloud Computing & IoT: 5G Focused IoT with Cloud Solutions,” *Int. J. AI, BigData, Comput. Manag. Stud.*, vol. 6, no. 3, pp. 21–25, 2025, doi: 10.63282/3050-9416.IJAIBDCMS-V6I3P103.

[3] S. H. H. Madni, M. S. A. Latiff, Y. Coulibaly, and S. M. Abdulhamid, “Recent advancements in resource allocation techniques for cloud computing environment: a systematic review,” *Cluster Comput.*, 2017, doi: 10.1007/s10586-016-0684-4.

[4] H. singh B. P. Singh, “Using LLMs for Autonomous Cloud Infrastructure Entitlement Management to Prevent Overprivileged Access,” *Sarcouncil J. Eng. Comput. Sci.*, vol. 1, 2026, [Online]. Available: https://doi.org/10.5281/zenodo.19488212

[5] M. Ali, S. U. Khan, and A. V. Vasilakos, “Security in cloud computing: Opportunities and challenges,” *Inf. Sci. (Ny).*, 2015, doi: 10.1016/j.ins.2015.01.025.

[6] V. K. Bollu, “Threat Landscape in Artificial Intelligence Systems: Taxonomy, Attack Vectors and Security Implications,” *World J. Adv. Res. Rev.*, vol. 29, no. 1, pp. 285–294, 2026, doi: 10.30574/wjarr.2026.29.1.0007.

[7] S. Chatterjee, “A Data Governance Framework for Big Data Pipelines: Integrating Privacy, Security, and Quality in Multitenant Cloud Environments,” *Tech. Int. J. Eng. Res.*, vol. 10, no. 5, 2023, doi:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, June 2026 Impact Factor: 8.2

10.56975/tijer.v10i5.158181.

[8] D. Bhattacharjee, "Design and Evaluation of Deep Generative AI Model for Intrusion Detection in Cyber Threat Monitoring," in *2025 7th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, IEEE, Dec. 2025, pp. 1–6. doi: 10.1109/ISAECT68904.2025.11318752.

[9] M. Thangavel, P. Varalakshmi, and S. Sridhar, "An analysis of privacy preservation schemes in cloud computing," in *Proceedings of 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016*, 2016. doi: 10.1109/ICETECH.2016.7569230.

[10] M. Armbrust, A. Fox, and R. Griffith, "Above the clouds: A Berkeley view of cloud computing," *Univ. California, Berkeley, Tech. Rep. UCB*, 2009, doi: 10.1145/1721654.1721672.

[11] S. Slimani, T. Hamrouni, and F. Ben Charrada, "Service-oriented replication strategies for improving quality-of-service in cloud computing: a survey," *Cluster Comput.*, 2021, doi: 10.1007/s10586-020-03108-z.

[12] V. I. C. Chang, "A Proposed Framework for Cloud Computing adoption," *Int. J. Organ. Collect. Intell.*, 2016, doi: 10.4018/ijoci.2016070105.

[13] P. Mell and T. Grance, "The NIST definition of cloud computing," in *Cloud Computing and Government: Background, Benefits, Risks*, 2011. doi: 10.1016/b978-0-12-804018-8.15003-x.

[14] N. Mostafa, H. S. M. Ramadan, and O. Elfarouk, "Renewable energy management in smart grids by using big data analytics and machine learning," *Mach. Learn. with Appl.*, 2022, doi: 10.1016/j.mlwa.2022.100363.

[15] S. R. Chanthati, "Leveraging Artificial Intelligence for Smart Cloud Migration, Reducing Cost, and Enhancing Efficiency," in *Smart Technology and Artificial Intelligence (STAI 2026)*, Zenodo, Apr. 2026. doi: 10.5281/zenodo.20319019.

[16] H. N. Dholariya, "Human-in-the-Loop AI for Cloud Data Engineering: The Collaborative Intelligence Architecture (CIRA) for Regulated Industries," *J. Inf. Syst. Eng. Manag.*, vol. 11, no. 1, 2026.

[17] S. M. Shivam, P. V. Anupama, and P. Jayachandran, "Anomaly Detection in Cloud Networks using Machine Learning Techniques," in *Proceedings of 3rd International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2025*, 2025. doi: 10.1109/ICSCDS65426.2025.11167399.

[18] M. Mushtaq, U. Akram, I. Khan, S. Khan, A. Shahzad, and A. Ulah, "Cloud Computing Environment and Security Challenges: A Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, pp. 183–195, 2017, doi: 10.14569/IJACSA.2017.081025.

[19] P. Kaur and M. Sachdeva, "A Survey On Cloud Computing and Its Various Services," *Int. J. Comput. Technol.*, vol. 15, no. 3, pp. 6626–6631, Jan. 2016, doi: 10.24297/ijct.v15i3.1687.

[20] L. Wang, J. Tao, M. Kunze, A. C. Castellanos, D. Kramer, and W. Karl, "Scientific cloud computing: Early definition and experience," in *Proceedings - 10th IEEE International Conference on High Performance Computing and Communications, HPCC 2008*, 2008. doi: 10.1109/HPCC.2008.38.

[21] A. Bakshi and B. Yogesh, "Securing cloud from DDOS attacks using intrusion detection system in virtual machine," in *2nd International Conference on Communication Software and Networks, ICCSN 2010*, 2010. doi: 10.1109/ICCSN.2010.56.

[22] L. Youseff, M. Butrico, and D. Da Silva, "Toward a unified ontology of cloud computing," in *Grid Computing Environments Workshop, GCE 2008*, 2008. doi: 10.1109/GCE.2008.4738443.

[23] J. Governor, D. Hinchcliffe, and D. Nickull, *Web 2.0 Architectures: What entrepreneurs and information architects need to know*. 2009.

[24] S. Murugesan, "Understanding Web 2.0," *IT Prof.*, 2007, doi: 10.1109/MITP.2007.78.

[25] R. Bhadauria and S. Sanyal, "Survey on Security Issues in Cloud Computing and Associated Mitigation Techniques," *Int. J. Comput. Appl.*, 2012, doi: 10.5120/7292-0578.

[26] J. H. P. Eloff, M. M. Eloff, M. T. Dlamini, and M. P. Zielinski, "Internet of People, Things and Services - The Convergence of Security, Trust and Privacy," *3rd CompanionAble Work. - IoPTS*, 2009.

[27] S. Ahmadi, "Zero Trust Architecture in Cloud Networks: Application, Challenges and Future Opportunities," *J.*

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 11, June 2026

Impact Factor: 8.2

Eng. Res. Reports, vol. 26, no. 2, pp. 215–228, 2024, doi: 10.9734/jerr/2024/v26i21083.

[28] Jericho Forum, "Cloud Cube Model: Selecting Cloud Formations for Secure Collaboration Problem," *Position Pap.*, 2009.

[29] J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, "Ensuring security and privacy preservation for cloud data services," *ACM Comput. Surv.*, 2017, doi: 10.1145/2906153.

[30] N. Mary and K. Jayapriya, "An Extensive Survey on QoS in Cloud Computing," *Int. J. Comput. Sci. Inf. Technol.*, 2014.

[31] A. Abdelmaboud, D. N. A. Jawawi, I. Ghani, A. Elsafi, and B. Kitchenham, "Quality of service approaches in cloud computing: A systematic mapping study," *J. Syst. Softw.*, 2015, doi: 10.1016/j.jss.2014.12.015.

[32] R. K. Gadiraju, "Artificial Intelligence for Resource Optimization in Cloud Computing Environments," *J. Electr. Syst.*, vol. 20, no. 6s, pp. 3164–3174, Apr. 2024, doi: 10.52783/jes.9485.

[33] T. P. Patel, S. Shivam, and V. Ranganathan, "Compute-Optimal Resource Allocation for Distributed Large Language Model Inference in Cloud-Scale Intelligent Systems," in *2026 14th International Symposium on Digital Forensics and Security (ISDFS)*, 2026, pp. 1–6. doi: 10.1109/ISDFS69419.2026.11458958.

[34] J. B. Mehta, "Autonomous Workload Right-Sizing for Multi-Cloud Cost Optimization," in *2026 International Conference on Artificial Intelligence, Systems, and Emerging Technologies (ICAISET)*, IEEE, Apr. 2026, pp. 1–11. doi: 10.1109/ICAISET66439.2026.11541899.

[35] D. Ardagna, G. Casale, M. Ciavotta, J. F. Pérez, and W. Wang, "Quality-of-service in cloud computing: modeling techniques and their applications," *J. Internet Serv. Appl.*, vol. 5, no. 1, p. 11, Dec. 2014, doi: 10.1186/s13174-014-0011-3.

[36] R. Nallabollu, D. S. Brunda, K. Swathi, P. Shabana, S. M. Nadeem, and T. Kavitha, "Parameter Efficient Fine Tuning of Large Language Models for Cloud Based Data in Low Resource Environments," in *2026 Contemporary Computing Innovations Conference (CCIC)*, IEEE, Feb. 2026, pp. 1–5. doi: 10.1109/CCIC68129.2026.11485947.

[37] Y. Suthari, Z. Thakker, S. Govindarajan, N. Krishnan, V. Raju, and S. K. Rao Katta, "Cost Optimization Techniques for Efficient Resource Allocation in Cloud Computing Environments," 2025. doi: 10.1109/icoici65217.2025.11253514.

[38] P. Vij and A. Mishra, "AI-Driven Resource Optimization in Cloud and Edge Environments for Sustainable Ubiquitous Computing," in *2025 5th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, IEEE, Nov. 2025, pp. 458–462. doi: 10.1109/ICUIS67429.2025.11380770.

[39] S. Hassan, Q. Li, M. Hassan, A. Yasin, M. Zubair, and M. Umair, "Unveiling the Application of Artificial Intelligence in Cloud Computing Environment," in *2024 5th International Conference on Innovative Computing, ICIC 2024 - Proceedings*, 2024. doi: 10.1109/ICIC63915.2024.11116294.

[40] P. S and R. S. Ponmagal, "Trust based Reputation Framework for Data Center Security in Cloud Computing Environment," in *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Feb. 2023, pp. 1041–1047. doi: 10.1109/ICCMC56507.2023.10084093.

[41] A. Karabetian *et al.*, "An Environmentally-sustainable Dimensioning Workbench towards Dynamic Resource Allocation in Cloud-computing Environments," in *13th International Conference on Information, Intelligence, Systems and Applications, IISA 2022*, 2022. doi: 10.1109/IISA56318.2022.9904367.