ADVANCED METHODS FOR MAINTAINING AND MANAGING THE LIFE CYCLE OF CLOUD ENVIRONMENTS: SURVEY

Resource management is a fundamental concept in cloud computing and virtualization, encompassing the allocation, release, coordination, and monitoring of cloud resources to optimize efficiency. The complexity arises from the virtualized, heterogeneous, and multi-user nature of these resources. Effective governance is challenging due to uncertainty, large-scale infrastructures, and unpredictable user states. This paper presents a comprehensive taxonomy of resource management technologies, offering a detailed analysis of design architecture, virtualization, and cloud deployment models, along with capabilities, objectives, methods, and mechanisms. In a cloud computing environment, deploying application-based resource management techniques necessitates understanding the system architecture and deployment model. This paper explores centralized and distributed resource management system architectures, providing a review of effective resource management techniques for both, accompanied by a comparative analysis. The evolution of cloud computing from a centralized to a distributed paradigm is examined, emphasizing the shift towards distributed cloud architectures to harness the computing power of smart connected devices at the network edge. These architectures address challenges like latency, energy consumption, and security, crucial for IoT-based applications. The literature proposes various methods for distributed resource management, aligning with the distributed nature of these architectures. Resource management in cloud computing involves discovery, provisioning, allocation, and monitoring functions, with sub-functions like mapping and scheduling. Integrated approaches to consolidation and resource management have been explored in numerous studies. This paper summarizes and analyzes existing research on resource management functions, focusing on identification, provisioning, allocation planning, and monitoring, based on their objectives and methods.

Keywords: cloud, cloud environment, life cycle of cloud environment, information technology

1 Introduction

In the ever-evolving landscape of Information Technology (IT), the advent of cloud computing has emerged as a transformative force, redefining the way organizations conceptualize, deploy, and manage their computing resources. Cloud environments, characterized by on-demand access to a shared pool of configurable computing resources, have become indispensable components of modern IT infrastructure. This paradigm shift has not only revolutionized the traditional IT landscape but has also introduced new possibilities and challenges for businesses and technology professionals [1].

1.1 Contextualizing Cloud Environments

Traditional IT infrastructure often grappled with issues such as resource constraints, high upfront costs, and limited scalability. The introduction of cloud environments has alleviated these challenges by offering a flexible and scalable model for provisioning computing resources. Organizations can now harness the power of virtualization and distributed computing, enabling them to scale their operations dynamically in response to changing demands [2].
1.2 Driving Forces of Cloud Adoption
The adoption of cloud environments is driven by a confluence of factors, including the need for cost efficiency, improved agility, and enhanced resource utilization. Cloud computing allows organizations to optimize their IT expenditure by shifting from a capital-intensive model to a more agile operational expenditure approach. This flexibility not only reduces financial barriers to entry but also enables businesses to adapt swiftly to market changes and innovation [3].

1.3 Enabling Technological Advancements
Cloud environments serve as the foundation for a myriad of technological advancements, including the proliferation of Big Data analytics, Artificial Intelligence (AI), and Internet of Things (IoT) applications. The scalable and elastic nature of cloud resources facilitates the seamless integration of these cutting-edge technologies, empowering organizations to derive actionable insights, enhance decision-making processes, and gain a competitive edge in the digital era [4].

1.4 Challenges in Cloud Environment Management
While the benefits of cloud computing are undeniable, the complexity of managing cloud environments throughout their life cycle poses challenges. Efficiently navigating the planning, deployment, monitoring, security, and maintenance phases requires robust methodologies and tools. Addressing these challenges is crucial for organizations seeking to maximize the potential of cloud environments while ensuring reliability, security, and compliance [5].

In light of these considerations, this paper explores advanced methods for maintaining and managing the life cycle of cloud environments. By addressing the complexities associated with cloud management, this research aims to contribute to the ongoing discourse on optimizing the utilization of cloud resources and ensuring the sustained efficiency and security of modern IT infrastructure.

2 Challenges of Managing the Life Cycle of Cloud Environments
The dynamic nature of cloud environments introduces a set of intricate challenges that organizations must navigate to ensure the optimal performance, security, and efficiency of their IT infrastructure. Understanding and addressing these challenges are imperative for successful cloud management. In this section, we delve into key challenges associated with various phases of the cloud life cycle [6].

2.1 Planning Phase Challenges
The planning phase sets the foundation for effective cloud management, yet it is fraught with challenges. Organizations often face difficulties in accurately estimating resource requirements, leading to over-provisioning or under-provisioning. Additionally, aligning cloud strategies with business goals while considering cost implications poses a constant challenge.

2.2 Deployment Phase Challenges
Efficient deployment of applications and services in the cloud is essential for minimizing downtime and ensuring a seamless transition. However, challenges such as compatibility issues, data migration complexities, and the need for robust automation tools often hinder the smooth deployment of resources. Balancing the speed of deployment with the need for thorough testing further complicates this phase.

2.3 Monitoring and Optimization Phase Challenges
Real-time monitoring and continuous optimization are critical for maintaining the performance and cost-effectiveness of cloud environments. However, the sheer volume and diversity of data generated by cloud resources can overwhelm traditional monitoring systems. Achieving a balance between resource optimization and ensuring high availability poses a persistent challenge.

2.4 Security and Compliance Phase Challenges
Security remains a paramount concern in cloud environments. Ensuring data confidentiality, integrity, and availability while complying with industry regulations and standards is a multifaceted challenge. The dynamic nature of the cloud, coupled with the shared responsibility model, necessitates robust security measures, yet implementing and maintaining these measures can be complex.

2.5 Maintenance and Upgrades Phase Challenges
Routine maintenance and upgrades are essential for keeping cloud environments secure and up-to-date. However, organizations often grapple with challenges related to scheduling maintenance windows that minimize disruption, testing compatibility with existing applications, and implementing updates seamlessly across diverse cloud resources.

2.6 Cross-Phase Challenges
Moreover, challenges often transcend individual life cycle phases. Coordinating activities across planning, deployment, monitoring, security, and maintenance while maintaining an agile and responsive infrastructure poses a systemic challenge. Addressing these challenges requires advanced methodologies, tools, and best practices. The subsequent sections of this paper delve into innovative methods designed to overcome these hurdles, aiming to contribute to the ongoing discourse on effective cloud environment life cycle management.
3 Related works

Resource management is one of the core concepts of cloud computing and virtualization. It is the process of allocating, releasing, coordinating, and monitoring cloud resources, enabling cloud systems to use cloud resources efficiently and effectively. Cloud computing relies on virtualization technologies, and such resources make cloud systems more complex because the resources are virtualized, heterogeneous, and multi-user. Uncertainty, heterogeneity, large cloud infrastructures, and the unpredictable state of the system due to the large number of cloud users make effective governance management very difficult. The large scale makes it difficult to obtain accurate and reliable global state information, and the large number of users makes it difficult to predict the type and intensity of workloads on cloud systems. Cloud RM requires autonomous, resilient, and efficient strategies that can avoid over-provisioning and under-provisioning of cloud resources and improve resource utilization. This section presents a taxonomy of existing resource management technologies and a detailed analysis, design architecture, virtualization, and cloud deployment models, capabilities, objectives, methods, and mechanisms based on them.

3 Architecture

In a cloud computing environment, before designing application-based resource management techniques and adapting them to the target computing environment, it is important to first consider the deployment model and system architecture of the environment. This subsection briefly describes centralized and distributed resource management system architectures and reviews some existing effective RM techniques for both architectures, along with a comparative analysis of centralized and distributed RM techniques.

3.1 Centralized Resource Management Techniques

Cloud computing has emerged as a new paradigm that has changed the computing paradigm from a distributed client-server architecture model to a centralized architecture model. Overall, it changed the era of computing from one in which basic computing resources were scarce and expensive to one in which the same resources were available and affordable. Traditional cloud systems concentrated cloud resources in centralized data centers with virtually unlimited capacity. The modern approach to cloud computing is based on dedicated data centers managed by the enterprise, where resources are perceived as unlimited and everything is delivered as a service in a fixed resource setting. Cloud computing democratized computing. It has created the illusion of unlimited computing, has jump-started the commercialization of computing, and has made the concept of utility computing a reality. Existing cloud computing developments have emerged as part of a centralized paradigm that concentrates available computing power in large, well-equipped data centers.

A number of centralized resource management techniques have been proposed in the literature to support cloud applications, employing a centralized architecture.

In [7] an agent-based elastic task package cloud, which is a centralized model was developed.

In [8] a multi-agent model, which is partially centralized by separate clusters of centralized resource provisioning mechanisms was developed. Multiple agents make cloud computing more flexible and more autonomous.

[9] presents integrate various software-defined system elements into an MEC system to provide an efficient software-defined model that provides mobile cloud computing services.

The problem of energy-efficient data offloading was studied in [10], where computational offloading is jointly optimized along with radio resource allocation to all users in the network to minimize energy consumption under delay conditions.

The paper [11] presents the problem of assuming personalized delay requirements of mobile users.

In [12] a multi-level control framework for resource allocation and load balancing to ensure QoS and minimize energy consumption is presented.

In [13] investigated allocation and tolerance management issues, which aim to minimize response times and optimize the use of computing resources on edge servers was studied.

[14] presents a scalable edge computing framework for early fire detection using image processing services hosted on an edge cloud. A scalable edge computing framework is proposed, where the application response time control is based on linear optimization and state feedback controllers.

In [15] an optimal multi-server MEC environment with multiple end-users was proposes. It presented the joint problem of end-user selection of mobile cloud computing (MEC) servers with data offloading and optimal pricing by MEC servers.

In [16] a new algorithm to improve intelligent dynamics (dynamic switching algorithm) for intelligent offloading of computing in edge clouds was developed.

In [17] an orchestration framework for intelligent offloading of computing in edge clouds was presented.

3.1.2 Distributed Management of Hardware Resources

Cloud computing has evolved as a centralized paradigm for creating pools of virtualized computing resources and making them available on demand to cloud users. However, the ever-increasing computing power of smart connected things and devices requires distributed clouds to avoid unnecessary latency and fully utilize the computing power available at the network edge. These distributed cloud architectures represent a major
transformation from a cloud perspective, but build on established research areas such as mobile cloud computing, mobile ad hoc computing, and fog/edge computing. These new distributed technologies overcome most of the challenges faced by centralized cloud computing data centers, such as latency, energy consumption, and security, which are critical issues for sensor-driven applications to support IoT-based applications. Many methods for distributed resource management, adapting distributed architectures, have been proposed in the literature.

In [18] an applied economic theory and pricing mechanism for offloading in mobile cloud systems that maximizes overall system benefits was proposed.

In [19] a computational a multi-level cloud-based offloading framework for offloading is proposed, which aims to maximize the utility function of both the vehicle and the computing server was proposed.

The paper [20] proposed an IoT-enabled device in the MEC era with They study the problem of QoS fulfillment based on game-theoretic learning: autonomous activation of MEC servers can be formulated as a few games, and a distributed learning approach can decide whether to activate or not.

[21] study the problem of computational offloading at mobile edge/fog based on autonomous and deep Q learning They study the problem of offloading computation in It aims to minimize latency and energy consumption.

The data offloading problem is studied in [22] by jointly optimizing computation offloading along with radio resource allocation to all network users to minimize energy consumption under completion time constraints.

Paper [23] studied a green hybrid greedy maximal scheduling method to solve the computation offloading problem in mobile edge clouds to improve performance and efficiency. To solve the computational offloading problem.

Authors [24] proposed multiple greedy maximal scheduling methods, which improve resource utilization, minimize total cost and delay, reduce handoff delay during offloading, and reduce the communication resource in 5G C-RAN and computing resource optimization problems.

In [25] multiple resource allocation to maximize the profit of mobile service providers was proposed.

In [26] a method to maximize the profit of mobile service providers by combining the network resources of the radio access cloud with the mobile cloud's proposed a unified framework for jointly scheduling computing resources. Ensure that devices meet minimum quality of service (QoS) requirements was proposed.

In [27] considered the risk-averse or loss-averse behavior of users in the final decision regarding the portion of computation tasks to offload to each server in a multi-MEC server environment. The study was based on the following principles was developed.

In [28] the concept of auction theory, proposed a cooperative resource allocation method that considers distributed resource allocation based on bidirectional auctions in which small base stations and users participate jointly. It aims to minimize the average response time; Zhao et al. [29] jointly address the problem of optimizing computational offloading and resource allocation by considering system utility and computation time.

In [29] an artificial intelligence-based framework to solve the system energy consumption and time delay was proposed. Authors studied the offloading problem to minimize it.

In [30] a technique to address the problem of mapping IoT applications at the edge and dynamically migrating parts of the application to meet SLAs. From fully centralized to fully distributed was proposed.

In [31] resource allocation and provisioning algorithms and evaluated them under different system parameters in terms of CPU utilization and query deadline compliance was proposed. Authors demonstrated that a distributed strategy that does not require coordination among computing points achieves performance close to the optimal fully centralized strategy with coordination overhead.

### 3.2 Resource Management Functions

Resource management in cloud computing has four functions: discovery, provisioning, allocation, and monitoring. In addition, there are sub-functions such as mapping and scheduling, but since these functions and resource allocation are specialized concepts, the RM model attempts to treat them collectively. Many studies have developed integrated approaches to consolidation and resource management. This section summarizes and analyzes existing research on these resource management functions (identification, provisioning, allocation planning, and monitoring) based on their objectives and the methods used.

In [32] a dynamic SLA-driven cloud RM to reduce execution time, but it has limitations for large computations was proposed.

In [33] an energy efficiency and SLA efficiency, and presented a multi-criteria RM heuristic was proposed. However, this model increases resource use for secondary tasks.

In [34] a cloud broker approach based on a genetic algorithm with biased random keys for RM in multi-clouds to improve efficiency at low cost was developed.

In the paper [35] an intelligent RM based on reinforcement learning for blockchain-based cloud data centers, achieving lower cost and execution time was developed.

In [36] a dynamic cloud-based RM for media applications with better QoS and maximum benefit. However, this model has data offloading limitations was developed.
### Table 1. Approaches’ analysis

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Framework base</th>
<th>Approach scheme</th>
<th>Related works</th>
<th>Research components</th>
<th>Approach via</th>
<th>Environment via</th>
<th>Deployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>[19]</td>
<td>Allocation of offloading and radio resources</td>
<td>+</td>
<td>+</td>
<td>Energy consumption</td>
<td>Improvement</td>
<td>Centralization</td>
<td>Edge computing for mobile devices</td>
</tr>
<tr>
<td>[9]</td>
<td>Management and control</td>
<td>+</td>
<td>+</td>
<td>Expense, latency</td>
<td>Special software</td>
<td>Centralization</td>
<td>Edge computing for mobile devices</td>
</tr>
<tr>
<td>[18]</td>
<td>Resource pricing</td>
<td>–</td>
<td>+</td>
<td>System maximization</td>
<td>Improvement</td>
<td>Distribution</td>
<td>Edge computing for mobile devices</td>
</tr>
<tr>
<td>[22]</td>
<td>Resource provisioning</td>
<td>+</td>
<td>+</td>
<td>Revenue maximization</td>
<td>Auction based approach</td>
<td>Distribution</td>
<td>Wireless networks</td>
</tr>
<tr>
<td>[11]</td>
<td>Offloading of computation, And allocation of radio and computing resources</td>
<td>+</td>
<td>+</td>
<td>Expense, latency</td>
<td>approximation</td>
<td>Centralization</td>
<td>Edge computing for mobile devices</td>
</tr>
<tr>
<td>[20]</td>
<td>Offloading of computation</td>
<td>+</td>
<td>+</td>
<td>Quality of Service, power usage</td>
<td>Machine learning informed by game theory</td>
<td>Distribution</td>
<td>Edge computing for mobile devices</td>
</tr>
<tr>
<td>[12]</td>
<td>The load balancing and the allocation of the resource</td>
<td>–</td>
<td>+</td>
<td>Quality of Service, power usage</td>
<td>Control with fuzzy logic Takagi-Sugeno</td>
<td>Centralization</td>
<td>Cloud</td>
</tr>
<tr>
<td>[21]</td>
<td>Offloading of computation</td>
<td>–</td>
<td>+</td>
<td>Expense, latency</td>
<td>Deep Q-learning for autonomic computing</td>
<td>Distribution</td>
<td>Computing at the mobile edge/fog</td>
</tr>
<tr>
<td>[23]</td>
<td>Offloading of computation</td>
<td>+</td>
<td>+</td>
<td>Utility maximization</td>
<td>Improvement</td>
<td>Hybridization</td>
<td>Edge computing for mobile devices via green technologies</td>
</tr>
<tr>
<td>[15]</td>
<td>Data offloading</td>
<td>–</td>
<td>+</td>
<td>Profit maximization</td>
<td>Stochastic learning automata theory, game theory, optimization</td>
<td>Centralization</td>
<td>Edge computing for mobile devices</td>
</tr>
<tr>
<td>[26]</td>
<td>Data offloading with risk awareness and best allocation</td>
<td>+</td>
<td>+</td>
<td>Utility maximization</td>
<td>Optimization with convexity, considering probabilistic uncertainty</td>
<td>Distribution</td>
<td>Computing at the edge</td>
</tr>
<tr>
<td>[14]</td>
<td>Detection of fires, scaling both horizontally and vertically</td>
<td>+</td>
<td>+</td>
<td>Effectiveness, power usage</td>
<td>Linear optimization for image processing</td>
<td>Centralization</td>
<td>Computing at the edge</td>
</tr>
<tr>
<td>[28]</td>
<td>Offloading of computation and allocation of resources</td>
<td>–</td>
<td>+</td>
<td>System usage and timing</td>
<td>Improvement</td>
<td>Distribution</td>
<td>MEC for cloud computing</td>
</tr>
<tr>
<td>[17]</td>
<td>Energy consumption</td>
<td>–</td>
<td>+</td>
<td>Utility of the system and processing time</td>
<td>Network Function Virtualization (NFV) coordination</td>
<td>Centralization</td>
<td>Cloud at the edge</td>
</tr>
<tr>
<td>[29]</td>
<td>Offloading tasks and allocating resources</td>
<td>+</td>
<td>+</td>
<td>Expense, latency, energy consumption, timing</td>
<td>Reinforcement learning, Improvement</td>
<td>Distribution</td>
<td>Edge computing for mobile devices</td>
</tr>
<tr>
<td>[16]</td>
<td>Smart, dynamic offloading</td>
<td>+</td>
<td>+</td>
<td>Expense, latency and energy consumption</td>
<td>Classification, optimization</td>
<td>Centralization</td>
<td>Cloud at the edge</td>
</tr>
<tr>
<td>[31]</td>
<td>Allocation and provisioning of resources</td>
<td>–</td>
<td>+</td>
<td>CPU usage</td>
<td>approximation</td>
<td>Hybridization</td>
<td>Cloud at the edge</td>
</tr>
</tbody>
</table>
Conclusion

The primary objective of the paper was to explore methods for maintaining and managing the life cycle of cloud environments. The scope encompasses a comprehensive examination of challenges encountered across the various phases of the cloud life cycle and the development of innovative solutions to address these challenges. As the result of the research the key challenges faced by organizations in planning, deploying, monitoring, securing, and maintaining cloud environments were investigated.

The research also gave the answer that the present situation requires such improvements and solutions:

1. To develop advanced methods, propose innovative and advanced methods tailored to overcome the challenges identified in each phase of the cloud life cycle.
2. To evaluate effectiveness, conduct empirical studies, including case studies and experiments, to evaluate the effectiveness of the proposed methods in real-world scenarios.
3. To provide practical insights, offer practical insights and recommendations based on the findings, aiding organizations in the implementation of robust strategies for cloud environment life cycle management.
4. To implement the securing of the cloud environment, that must involve executing various measures to protect data, applications, and infrastructure hosted in the cloud [37-39].

References