Efficient and Scalable Resource Management in Cloud-Based IoT Environments: A Systematic Literature Review

Nosheen Qamar[1], Faria Nazir[1], Nosheen Sabahat[2], Sundas Sagheer [1], Maham Noor[3]

[1]Department of Software Engineering, University of Management & Technology, Lahore, Pakistan

[2]Department of Computer Science, Forman Christian College University, Lahore, Pakistan.

[3]Department of Software Engineering, University of Central Punjab, Lahore, Pakistan

***Correspondence**: nosheen.qamar@umt.edu.pk

Citation | Qamar. N, Nazir. F, Sabahat. N, Sagheer. S, Noor. M, “Efficient and Scalable Resource Management in Cloud-Based IoT Environments: A Systematic Literature Review”, IJIST, Vol. 8 Issue. 2 pp 843-867, May 2026

Received | March 24, 2026 **Revised |** April 26, 2026 **Accepted |** May 01, 2026 **Published |** May 07, 2026.

Introduction: The increasing rate of spread of the Internet of Things (IoT) has led to a massive amount of data and poses a pivotal challenge in the resource-sharing process of a cloud computing environment. It is needed to manage resources effectively, with scalability, to provide support for IoT operations at reasonable costs and reliability.

Novelty Statement: This Systematic Literature Review (SLR) examines literature on resource optimization techniques in IoT-enabled cloud environments with a focus on reducing resource consumption, yet system performance continues to be maintained.

Material and Method: This study examines the related research on prominent scholarly databases depending on the adaptation of dynamic resource provision, workload balancing, cost-based provisioning, and proactive scaling strategies. It also identified machine learning-based demand forecasting, adaptive resource scaling approaches, and containerization as efficient deployment methods. Another aspect that is mentioned in the review is the importance of the analysis of historical data and the intelligent distribution of workload for the improvement of scalability and responsiveness.

Result and Discussion: The results indicate an increasing interest in the flexible and elastic resource management dimensions that address the changes in workloads and varying application demands. This review provides discussion on sustainable and efficient cloud-based approaches to IoT since it summarizes the most current solutions and creates a research gap through the identifies research gaps and opportunities. Findings from reviewed studies indicate improvements, including 17.3% faster execution time, 22% higher throughput, 18% energy savings, 23% lower provisioning cost, and 87% resource efficiency in cloud-based IoT environments. The results indicate that security vulnerabilities have the highest impact severity score (10), while rule-based scaling and infrastructure control-related challenges show the lowest scores (6–7). Furthermore, auto-scaling demonstrates the highest effectiveness score (10) among dynamic scaling strategies, whereas rule-based and vertical scaling exhibit comparatively lower effectiveness scores (6–7) in cloud-based IoT environments.

Conclusion Remarks: It underlines that there is a demand of optimized, intelligent systems that correlate performances with energy consumption and power management in medium- to large-scale IoT environments.

Keywords: Cloud-based IoT; Resource Management; Cost-Efficiency; Scalability; and Performance Optimization

Introduction:

In modern computing, however, the conjunction of cloud technologies with the Internet of Things (IoT) introduces a new digital paradigm, where data is generated, processed, and consumed. Today, cloud-based systems are known for their seamless integration within IoT devices and the cloud. They represent the basis of current information technologies. This convergence is like an accelerator for organizations utilizing the huge capacity of data from IoT; at the same time, they can leverage the distributed and flexible and scalable nature as well as its cost-effectiveness. A key feature of cloud-based IoT is that it allows the flow of data between the internet-connected devices and the cloud servers in a hassle-free manner [1]. This mechanism gives rise to real-time data and insights, prediction, and intelligent decision-making. By employing cloud resources, the businesses are able to step beyond the bounds of on-premises infrastructure and thus scale resources appropriately, taking into account the fluctuation of the load and based on the needs of the prevailing time. Moreover, the cloud gives an opportunity for ideas in any field of industry, for instance, e-commerce, to spread across a wider share of the population [2].

The effective management of cloud computing resources, such as computational, storage, and networking resources, is is at the core of cloud IoT deployments. A critical factor to ensure effective and scalable resource management is implementing cost-efficient cloud-based IoT system operations [3]. There is a need for deep analysis to make use of available resources optimally while keeping operating costs minimal. Organizations actively leveraging IoT investments to get the best from their IoT investments are slowly realizing the essence of proper management of resources on their way to ensuring the agility, operational efficiency, and innovation in the building of competitiveness [4].

Research Gap:

Although cloud-based IoT systems are increasingly being used, there remains a lack of detailed and systematic studies that concentrate on efficient and scalable resource management in those systems [3][4]. As far as we know, there are only a few studies that have investigated, in a comprehensive manner, the impact of resource allocation strategies, dynamic scaling mechanisms, workload balancing approaches, serverless computing, and cost optimization techniques in cloud-based IoT infrastructures [5]. Current research focuses on individual issues such as security, latency, and energy efficiency, and there is a lack of attention on integrated and scalable resource management frameworks that satisfy the requirements of performance, scalability, cost-effectiveness, and reliability [6]. Moreover, there is a lack of comparative and empirical analyses critically assessing the effectiveness, limitations and industrial applicability of prevailing approaches to resource management in real-world IoT-cloud deployments [7][8][9][10][11].

Research Objective:

The main goal of this research is to comprehensively explore efficient, cost-effective and scalable resource management strategies in IoT based cloud computing. The objectives of the study are to explore, identify and analyze the challenges, strategies, tools and optimization techniques being employed for resource allocation, workload balancing, dynamic scaling and serverless computing in IoT-cloud systems. Besides, the existing approaches are assessed based on scalability, latency, throughput, energy efficiency, security and cost optimization and the gaps in the existing research are identified and future research directions are proposed. This work contributes by offering a comprehensive and integrated SLR work that covers multiple dimensions of the problem of managing resources in cloud-based IoT applications, which are often studied in isolation in the existing literature – such as scalability, performance optimization, security, and cost-efficiency. Moreover, this work makes a significant contribution in the field by comparing the current techniques, discussing pros, cons, real-world applications, and industrial viability to provide research, practice, and decision-makers with useful insights and guidance in cloud-IoT environments.

In the following sections, we continue with the inner exploration. We move from questions to the meaning of that and discuss the propositions. We turn to the literature review, present the findings, and finally describe the implications for practice and future research. Through the course, we hope to make a meaningful contribution to the immersion of perspective around efficient and scalable resource management in cloud IoT and act as a catalyst for a transformational change and an open source of opportunities in connected computing.

Related Work:

Distributed computing and the Internet of Things enable a paradigm shift in computation nowadays by providing a means of elastic resource access and linking physical devices to turn data into information [12]. Nevertheless, by efficient use of resources in the cloud-based IoT environments, significant challenges arise due to variable workloads of computing power and the variety of devices used [13]. Previous publications are concentrated on dynamically adjusting to the use of resource allocation by local scaling and elasticity strategies [5]. Server-less computing is also a way to abstract the infrastructure component, while eliminating the management burdens and providing flexibility for scaling [14]. Security can be defined as the most important factor that needs to be taken into account in cloud-based IoT environments, and the encryption and access control mechanisms can act as a safeguard mechanism. Therefore, help prevent loss of data confidentiality and integrity [14].

Table 1 presents a comparative analysis of recent studies focusing on efficient and scalable resource management approaches in cloud-based IoT environments. The table summarizes different architectures, performance metrics, tools, and experimental outcomes reported in the literature, highlighting improvements in latency reduction, scalability, energy efficiency, throughput, cost optimization, and resource utilization. Furthermore, the comparison demonstrates the growing adoption of edge–cloud integration, federated learning, AI-driven resource management, and dynamic scaling techniques for enhancing the performance and reliability of modern IoT-cloud infrastructures [5][14][15].

Cloud-based IoT has made major strides, especially in terms of addressing the challenges in resource management. However, this just makes the resource management even more complex than in the past. These implications have not been addressed, and this is the area that needs further study [5]. IoT load balancing is a crucial step in maintaining the stability of cloud-enabled IoT systems at a higher level. This implies that the data understanding should be improved significantly, including dynamic scaling, server-less computing, and security protection measures. Dynamic scaling not only attempts to adapt resources, but also their processing power to the unexpected needs according to the load. A large portion of resources is used inefficiently on resources that can quite rarely ensure satisfactory results; instead, dynamic scaling at the very end of the process will eliminate such costs [15][16].

On the other hand, Serverless computing abstracts infrastructure management and does away with resource management and configuration complications - just as the situation is with the alternative approaches [5]. Such strategies play a gateway to resource optimization and operation efficiency using real-time active resource monitoring and management approaches, key for cloud-based IoT deployments. Besides, even if the size of a particular management app is small, the dynamic scaling, server-less computing, and solid security are important [17]. To be effective, these approaches must address unknown vulnerabilities and risk becoming the new threats to the network. Avenues for future research are to improve current strategies, explore new approaches, and address the challenges that arise [18]. Table 2 explains the most related studies and compares them with this paper regarding their survey approach, quality assessment, and research framework.

Table 1. Comparative Table of Real Metrics

Ref	Study and year	Architecture / Approach	Metrics Evaluated	Metric Values	Tools / Platforms	Key Contributions
[12]	Kaur et al. (2023)	Edge–Cloud Hybrid	Latency	45–120 ms	Docker, Kubernetes	Adjustment of workload migration decreased the delay of end-to-end services.
[13]	Kim & Ko (2023)	Distributed Split Computing	Scalability	2× workload with 12% overhead	MQTT, EdgeX Foundry	Cooperative strategies of task offloading allow high tolerance of workload.
[18]	Velmurugadass et al. (2021)	Edge–Cloud IoT Infrastructure	Reliability	99.92% uptime	AWS EC2, Prometheus	Ensured a robust level of issues and fault-tolerance, and the trustworthy conveyance of data in industrial IoT.
[19]	Salunkhe et al. (2023)	Decentralized Learning in IoT	Energy Efficiency	15–20% lower consumption	Google Cloud, TensorFlow Lite	Incremental learning would allow optimal energy utilization of limited devices.
[20]	Rani et al. (2022)	Fog–Cloud for Smart Cities	Storage Overhead	10–13%	Hadoop, Apache NiFi	Dynamic replication of urban sensor data optimized storage.
[6]	Jeyaraj et al. (2023)	Multi-agent AI Resource Manager	Cost Efficiency	23% lower provisioning cost	AWS Lambda, AI Agents	The IoT-cloud minimized the cost of performance through dynamic active learning-based provisioning.
[21]	Fang et al. (2024)	IoT-Edge Task Scheduling	Execution Time	17.3% faster than baseline	Kubernetes, ONOS	Better real-time response by means of latency-consciousness scheduling algorithms.
[22]	Sethi & Verma (2024)	Federated Edge–Cloud Learning	Energy & Accuracy	18% energy saved, 92% model accuracy	PyTorch Mobile, Edge TPU	Cross-edge federated learning model to reduce power consumption.
[23]	Wang & Zhao (2025)	SDN-Based IoT Data Routing	Throughput & Latency	22% higher throughput, 35ms avg. latency	SDN/OpenFlow, Mininet-WiFi	Collaborative flow rerouting to improve vehicular IoT networks' QoS.
[24]	Alavi & Tariq (2025)	Cloud–Fog Hybrid for Healthcare IoT	Resource Utilization	87% resource efficiency	IBM Cloud, Node-RED	Context-based variable scaling of health monitoring sensors IoT.

Table 2. Summary of Related Studies on Cost-Effective and Scalability

References	**Focus of Survey**	**Quality Assessment**	**Research Framework**	**Cloud Computing and IOT Tools**	**Targeted Digital Repositories**
[1]	The survey focused on the security of the cloud computing and strong relation with IoT.	×	√	×	SYMMETRY-BASEL
[3]	The survey focused on the experiences and gains of the bandwidth and resources related to the cloud services and the IoT technologies.	√	√	×	ACM COMPUT SURV
[12]	The focus is on the relationship between the IoT and cloud computing.	√	×	√	Research Gate
[14]	Review of status of cloud computing and IoT in the industry.	√	√	√	IEEE Xplore

Systematic Literature Review Process:

This section describes the proposed methodology used for the SLR. We used the [25] methods for the proposed SLR. The methodology employed in this systematic literature review is characterized by its robustness, rigor, and clarity of approach. To ensure comprehensive coverage of the research domain, a meticulous search strategy was devised, encompassing a wide range of reputable digital repositories, including IEEE Xplore, ACM Digital Library, Springer Link, and Science Direct. This strategy was informed by carefully selected keywords and search terms tailored to capture the multidimensional aspects of cost-effective resource management and scalability in cloud-based IoT environments.

Research SLR Planning/ Protocol:

To maintain objectivity and relevance, strict inclusion and exclusion criteria were established, guided by predefined research questions and objectives. Screening of identified studies was conducted in multiple stages, involving independent assessment by multiple reviewers to ensure consistency and accuracy in study selection.

Research Planning the Research Questions:

The culmination of this methodological rigor is a comprehensive synthesis of the current state-of-the-art knowledge in cost-effective and scalable resource management within cloud-based IoT contexts, providing valuable insights and directions for future research and practice in the field. The systematic approach of combining empirical studies, simulations, and a literature review ensures the robustness of conclusions and provides valuable insights and practical recommendations for optimizing resource management in cloud-based IoT environments.

RQ1: What are the challenges and opportunities associated with implementing server-less computing for resource management in cloud-based IoT, and how does it impact overall performance and security?

RQ2: What role do dynamic scaling strategies play in achieving cost-efficient resource utilization in cloud-based IoT environments, and how can these strategies be optimized for different application scenarios?

Research Inclusion and Exclusion Criteria:

To maintain objectivity and relevance, strict inclusion and exclusion criteria were established, guided by predefined research questions and objectives. Screening of identified studies was conducted in multiple stages, involving independent assessment by multiple reviewers to ensure consistency and accuracy in study selection.

Figure 1 shows the details from identification to inclusion/exclusion based on the PRISMA framework, which includes the identification of papers using databases, resulting in 166749 identified papers.

Identification Phase:

The identification stage involves searching and collecting research studies related to the subject from different scholarly databases and digital libraries. For this study, research papers on efficient and scalable resource management in cloud-based IoT environments were retrieved from various databases such as IEEE Xplore, ACM Digital Library, SpringerLink and other indexed databases. A total of 166,749 records were initially found by searching for predefined keywords and strings. At this stage, research studies that were obviously not related to the research topic were eliminated, and 123,234 studies were discarded.

Screening Phase:

The remaining studies were screened for eligibility in the screening stage by determining the relevance of the research from the title and abstract. The goal of this stage was to remove duplicate, irrelevant, and poor-quality studies that did not directly involve cloud-based IoT resource management, scalability, dynamic scaling or serverless computing.

43,506 records were screened for title and abstract, 36,510 studies were excluded on title and abstract screening, and 6,745 studies were deemed out of scope.

Eligibility Phase:

The eligibility stage comprises full-text analysis of selected studies, revealing if the studies meet the predefined inclusion and exclusion criteria. This study focused on analysing carefully the introduction, methodology, findings and discussion section of the papers. Only studies providing empirical or methodological contributions, provided empirical evidence, had a discussion of scalability, and were relevant to resource management in cloud-based IoT were included. A total of 251 studies were screened in detail, and 91 records were excluded and 93 studies were removed because the objectives of the systematic review did not match the aim of the screening study.

Synthesis (Inclusion) Phase:

The synthesis phase is the last step of the selection of studies for the systematic literature review. Following all of the filtering and quality assessment processes, 56 high-quality and relevant studies were selected for the final filtering and analysis and synthesis. The challenges and opportunities were systematically evaluated for cloud-based IoT environments in these studies, as well as scaling strategies, serverless computing approaches, performance optimization techniques, security concerns and cost-efficient resource management solutions. The results of these selected studies guided the answer to the research questions and the direction of future research.

To ensure the relevance and quality of selected literature, the study selection process was carried out by applying pre-defined inclusion/exclusion criteria. All published studies from 2019 to 2025 were considered, to reflect recent research advances for cloud-based IoT resource management. The following restrictions were applied to the selection of the sources: Only English-language, peer-reviewed journal articles and conference papers from well-established databases like IEEE Xplore, ACM Digital Library, SpringerLink or ScienceDirect were considered. Studies of resource allocation, dynamic scaling, scalability, serverless computing, workload balancing, and cost optimization in cloud-based IoT environments were included in the empirical, qualitative, quantitative, and mixed-method studies. Studies, non-peer reviewed articles, editorials, short papers and studies which could not be clearly understood or did not have clear methodology or were not directly relevant to the research objective were excluded. Also, quality criteria: methodological contribution, empirical evidence and publication quality were used to guarantee the inclusion of high-quality studies in the final review.

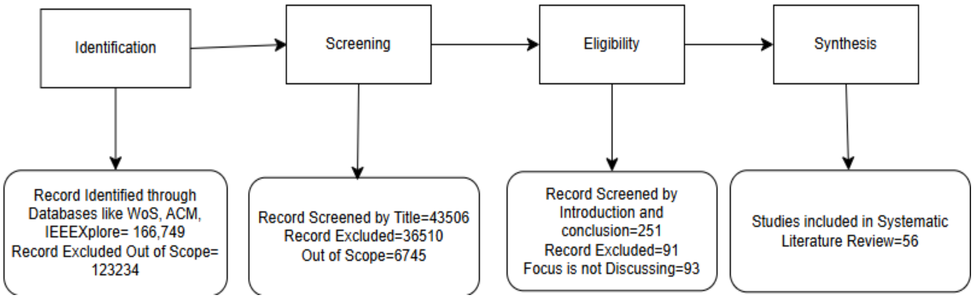

Figure 1. Inclusion/Exclusion Criteria

SLR Conduction:

To ensure comprehensive coverage of the research domain, a meticulous search strategy was devised, encompassing a wide range of reputable digital repositories, including IEEE Xplore, ACM Digital Library, SpringerLink, and ScienceDirect. This strategy was informed by carefully selected keywords and search terms tailored to capture the multidimensional aspects of cost-effective resource management and scalability in cloud-based IoT environments.

Data Gathering:

The search phase requires researchers to employ predetermined techniques across digital databases to locate relevant studies. The structured search protocol helps retrieve relevant scholarly and peer-reviewed articles related to the topic. The review process uses multiple online databases in accordance with predefined guidelines to retrieve scientific literature. Articles retrieved through the search process are filtered using predefined inclusion and exclusion criteria, and how can these strategies be optimized for different application scenarios?

Data Analysis:

Table 3. Quality Assessment

Ref	Research Type	Empirical Type	Methodology	(a)	(b)	(c)	(d)	Score
[1]	Adversary Model	Qualitative	Observational	0	1	0	4	5
[2]	IOT Communication Models	Qualitative	Interviews	1	0	1	3	5
[3]	request-reply interaction model, mF2C, spring	Mix Method	Experimental	1	1	2	4	8
[4]	IOT, IoTHeF	Quantitative	Interviews	1	1	1	4	7
[26]	No	Qualitative	Ground Theory	0	1	2	4	7
[15]	No	Quantitative	Interviews	0	1	2	4	7
[16]	Batch-oriented processing, Hadoop MapReduce, analytical, micro-batch, stream-oriente	Mix Method	Experimental	1	1	2	4	8
[27]	No	Qualitative	Interviews	0	1	2	3	7
[28]	Conceptual, NIST conceptual reference, IoT reference	Mix Method	Observational	1	1	2	4	8
[15]	No	Quantitative	Interviews	0	1	1	4	6
[29]	A CoT based healthcare	Qualitative	Experimental	1	1	2	3	7
[30]	Multiple Computing Frameworks	Qualitative	Observational	1	1	1	3	6
[31]	Service	Quantitative	Interviews	1	1	2	4	8
[32]	Three-layer IoT architectural	Mix Method	Experimental	1	1	2	4	8
[33]	No	Quantitative	Experimental	0	1	2	4	7
[6]	NO	Qualitative	Experimental	0	1	2	2	5
[21]	$^{3}\$$ design of framework for agri-FORXG	Quantitative	Ground Theory	1	1	1	4	7
[34]	IoT-CCPHM	Quantitative		1	1	2	4	8
[35]	Trango and PikeOS	Mix Method	Observational	1	1	2	4	8
[36]	NO	Quantitative	Interviews	0	1	1	4	6
[37]	DNN	Quantitative	Interviews	1	1	2	3	7
[38]	NO	Qualitative	Experimental	0	1	1	4	6
[7]	Supply chain	Mix Method		1	1	2	4	8
[8]	NO	Qualitative	Observational	0	1	1	4	6
[30]	Conceptual framework	Quantitative	Experimental	1	1	1	4	7

[31]	request-reply interaction model	Quantitative	Experimental	1	1	2	3	7
[14]	Knative framework	Quantitative	Interviews	1	1	1	4	7
[39]	NO	Quantitative	Experimental	0	1	2	4	7
[20]	NO	Mix Method	Ground Theory	0	1	2	3	6
[34]	NO	Quantitative	Observational	0	1	1	4	6
[40]	DNN Framework	Quantitative	Interviews	1	1	2	4	8
[41]	Mf2c	Quantitative	Experimental	1	1	1	4	7
[9]	Conceptual	Qualitative	Experimental	1	1	1	4	7
[42]	Three layer	Quantitative	Interviews	1	1	2	2	6
[43]	Multi computing	Quantitative	Experimental	1	1	2	4	8
[44]	request-reply interaction model	Qualitative	Observational	1	1	1	4	7
[45]	NIST		Experimental	1	1	2	3	7
[10]	Adversary model	Quantitative	Interviews	1	1	1	3	6
[46]	IOTHeF	Quantitative	Experimental	1	1	2	4	8
[47]	Conceptual	Mix Method	Ground Theory	1	1	1	4	7
[48]	DNN Framework	Qualitative	Observational	1	1	1	4	7
[21]	Knative framework		Interviews	1	1	1	4	7
[49]	CCPHM and Tango	Mix Method	Observational	1	1	1	4	7
[50]	SAS model	Qualitative	Experimental	1	1	2	3	7
[51]	Conceptual Framework	Quantitative		1	1	1	4	7
[18]	MCS, MIG, MES	Qualitative	Experimental	1	1	2	4	8
[52]	Back-end data sharing objects	Mix Method	Observational	1	1	1	4	7
[6]	CloudSim framework, mixed reality framework, and conceptual framework	Qualitative	Observational	1	1	2	4	8
[13]	DNN, Head, Tail Models	Quantitative	Observational	1	1	1	4	7
[19]	System and task resolution models	Mix Method	Interviews	1	1	2	2	6
[49]	Latency-aware task scheduling (Edge–IoT)	Quantitative	Experimental	1	1	2	4	8
[22]	Federated Edge–Cloud Learning (Healthcare)	Mix Method	Experimental	1	1	2	4	8
[23]	SDN-based Vehicular IoT Routing	Quantitative	Experimental	1	1	2	4	8
[53]	Fog–Cloud Hybrid Scaling (Healthcare)	Qualitative	Observational	1	1	2	3	7
[54]	Multi-Tenant Serverless Framework	Mix Method	Experimental	1	1	2	4	8
[11]	Edge AI with Reinforcement Learning	Quantitative	Experimental	1	1	2	4	8

The data analysis phase is classified into the following sub-sections.

Metadata Analysis: Quality Assessment and Publication year-wise: This section presents the quality assessment of studies using the following parameters:

Research framework: If the Results of the study define cost-effective and Scalable Resource Management in cloud-based IoT, for a Yes score (1); otherwise, for a No score (0).

Methodology: Score (2) for yes, novel and relevant contributions to Cost-Effective and Scalable Resource Management in Cloud-Based IoT, (1) for limited, and (0) for none.

Empirical type: Score (1) for Yes; if empirical evidence supports the Cost-Effective and Scalable Resource Management in Cloud-Based IoT; otherwise, score (0) for No.

Studies are evaluated based on the credibility of the publication sources, considering graded rankings of countries, journals, and conferences in computer science through their quartile ranking in JCR for journals and Core (A, B, C) for conferences. Table 3 explains the complete quality assessment parameters for our research area. The quality assessment columns represent: (a) Research Framework relevance, (b) Methodological contribution, (c) Empirical evidence support, and (d) Publication quality ranking.

Publication Source-Wise Metadata Analysis:

Table 4. Journals and No. of Publications

Sr No	Publication Source	No of Publications
1	Journal of Grid Computing	1
2	Soft Computing	2
3	IEEE Internet of Things Journal	1
4	Journal of Sensor and Actuator Networks	1
5	Journal of Ambient Intelligence and Humanized Computing	1
6	International Journal of Advanced Computer Science (Int J Adv Comput Sc)	3
7	Security and Communication Networks (Secure Commun Netw)	1
8	International Conference on Computational Performance Evaluation (ComPE)	1
9	10th International Symposium on Signal, Image, Video and Communications (ISIVC)	1
10	ACM Computing Surveys (ACM Comput Surv)	5
11	IEEE Access	3
12	IEEE Network	1
14	International Conference on Latest Developments in Materials and Manufacturing (ICLDMM)	1
16	Advances and Applications in Mathematical Sciences (Adv Appl Math Sci)	1
17	13th International Conference on Ubiquitous Information Management and Communication (IMCOM)	1
18	1st International Conference on Electronic Engineering and Renewable Energy (ICEERE)	1
19	16th Annual IEEE International Systems Conference (SysCon)	1
20	International Journal of Cloud Applications and Computing (Int J Cloud Appl Com)	1
21	29th International Conference on Computer Communications and Networks (ICCCN)	1
22	24th Optoelectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)	1

23	11th International Conference on Information and Communication Systems (ICICS)	1
24	3rd National Conference on Functional Materials (NCFM) - Emerging Technologies and Applications in Materials Science	1
25	16th IEEE International Colloquium on Signal Processing and its Applications (CSPA)	1
26	IEEE 9th International Conference on Cloud Networking (CloudNet)	1
27	IET communications	1
28	7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud) / 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)	1
29	Sensors (Basel)	2
30	Future Generation Computer Systems (Future Gener Comp Sy)	1
31	Internet of Things (Netherlands) (Internet Things-Neth)	1
32	2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)	1
33	Internet of things-based cloud computing platform for analyzing the physical health condition	1
34	2020 3RD INTERNATIONAL CONFERENCE ON INFORMATION AND COMPUTER TECHNOLOGIES (ICICT 2020)	1
35	7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud) / 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)	1
36	Future Internet	1
37	IETE Technical Review (IETE Tech Rev)	1
38	IEEE Transactions on Cloud Computing (IEEE T Cloud Computing)	1
39	3rd International Conference of Reliable Information and Communication Technology (IRICT)	1
40	IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)	1
41	Chinese Automation Congress (CAC)	1
42	Electronics (Switzerland)	1
43	3rd International Conference on Intelligent Computing in Data Sciences (ICDS)	1
44	SYMMETRY-BASEL	1

This includes the analysis of metadata, publication source wise source-wise. The publication sources include all the major journals and conferences that help in conducting this SLR. Table 4 demonstrates publication sources from which we have collected the papers for our SLR and the number of publications collected from each source.

Result and Discussion:

In this section, this section presents and discusses the findings from the different approaches proposed by the researchers

RQ1: What are the challenges and opportunities associated with implementing server-less computing for resource management in cloud-based IoT, and how does it impact overall performance and security?

Effectively implementing dynamic scaling strategies in cloud-based IoT environments is not without its challenges. Predictive scaling, while proactive, faces the hurdle of accurate predictions in dynamically changing environments with unpredictable workloads. Automation mechanisms may be very effective, but they need to be refined to find a balance between quick responses and cost-efficiency er. Horizontal scaling can cause system performance issues since such systems might develop bottlenecks in the path of distributed systems, which affect overall performance. By vertical scaling, one can increase efficiency at inter-instance level; however, vertical scaling alone cannot fully replace horizontal scaling. Hybrid scaling, the bidirectional embracement of both models, leads to that complexity must be resolved in running this model. Active scaling, that is, prompt reaction to fluctuations, is likely to result in obscene over-provisioning of resources when there is a load peak. Rule-based scaling with structural framework is intended to address various unexpected affairs, which are non-dynamic. Modern companies should be able to scaling costs-effectively amid real-time cost assessment with dynamic surroundings. Application-specific scaling requires trenchant work of profiling for successful implementation. This involves doing a lot for instance, identifying which issues merit scaling up and which other ones need to be scaled down. Time-based scaling which periodically adjusts the staff but probably does not allow for quick adaptation when suddenly workload changes. Overcoming these challenges demands this careful evaluation, involving an in-depth appreciation of unique features in each strategy, and accounting for the dynamic nature of IoT workloads, while ensuring the implemented strategy suits the specific needs of the application scenario.

Cold Start Latency:

Cold start up time (as in the case of function execution), will triggers the slowness of functions initiation, mostly because of dependent systems or real-time applications. It does not so much inhibit first-glance processes, but might interfere with user experience and efficiency

Resource constraints:

Survival of the fittest is one of the main barriers to confronting the computational, memory, and storage capacities problems. Scalability may be one of the beneficial features provided by server-less architecture; however, these constraints may affect the execution of difficult IoT tasks.

Vendor lock-in:

Applications tightly coupled with a cloud architecture can only be accessed through the given cloud service provider, usually complicating migration to a different cloud platform. Dependence on a vendor ecosystem may increase operational costs as such may make business costlier and less effective [42].

Monitoring and debugging:

On the other hand, overseeing system resources, and troubleshooting tasks have become more troublesome because of a lesser amount of visibility and control found in a server-less infrastructure. The availability of necessary tools and resources to support the issue identification and resolution process is scarce [48].

Security vulnerabilities:

Interdependence within the server-less framework and multi-tenancy necessitate protection of a large area, thus increasing the risk of potential security challenges. Although a well-thought-out security system has been designed in this case, offer involves exposure to data breaches and unauthorized access [44].

Performance unpredictability:

Time irregularity comes about through different operational cycles and availability of shared resource services. This parity can provide an unpredictable reaction under unpredictable conditions and responsiveness [46][47][48].

Scalability concerns:

Scalability problems that cause limits arise because of juggling multiple workloads, concurrency and resource distribution constraints within the Internet of Things. This shortcoming may, then, be detrimental to the organization's capacity to tackle heavy workloads productively.

Cost management complexities:

While costs might be challenging predicting and the managing factors are, more so from the emphasis on billing and usage specifics. However, managing expenses requires a balance of scenario that enables the optimal utilization of the resources including, the cost efficiency [13].

Integration challenges:

Integration of an incompatibility issues may harm when integrating server-less components with existing systems. It is important to make certain that connecting machinery efficiently is done by having clear communication and determining the compatibility levels.[46].

Lack of control over infrastructure:

Any reduced control over the underlying infrastructure means you have to forgo any customization and optimization options. The limitations in this context may rise and make it difficult to customize the environment for a given application

Network limitations:

Broadcast of the information across the provider networks can lead to the deterioration of the performance and latency problems. Overcoming network congestion is a very vital tasks in terms of obtaining the unbreakable performance of communication

Limited support for long-running tasks:

The spend of time for task completion is limited by obstructing prolonged computation projects. The given constraint can be broadened in terms of the methods that involve long running time and some of the applications may underperform due to the given constrain

Operational challenges:

Management overload has a potential perceptible growth of versioning, deployment, and monitoring complications. Cutting down on redundant processes is key for keeping stable the system functioning [44].

Debugging complexity:

Digital platforms' incremental upgrade structure, often event-driven systems, contribute to debugging problems. Debugging complexity necessitates the use of specific tools as well as relevant strategies that allow for thorough wearer-oriented problem-solving [9].

Limits on execution time and resources:

Resources and time limit impedes the ability to complete either critical path tasks or resources-demanding activities such as data processing, because they require high execution time. The applications that take up lots of resources find impeccable execution difficult [42]. Table 5 demonstrated the challenges and importance of associated with implementing server-less computing for resource management in cloud-based IoT. It also states the limitations of each study.

Table 5. Challenges, Importance, and Limitations of Studies

Reference Study	Challenges	Importance	Limitation
[14]	Cold start latency: This delay in function initialization affects real-time applications, potentially causing user experience issues, despite primarily affecting initial function invocations.	High: Delays real-time responsiveness, affecting user experience	Low: Mainly impacts initial function invocation
[5]	Resource constraints: Limitations on computation, memory, and storage hinder executing complex IoT tasks, despite server-less architectures offering scalability.	High: Affects execution of complex IoT tasks	Low: Scalability may mitigate some limitations
[42]	Vendor lock-in: Tying applications to specific providers limits flexibility, complicates migration, and may increase costs due to dependencies.	High: Limits flexibility and may increase costs	Low: Potential challenge when considering migration
[48]	Monitoring and debugging: Reduced visibility and tools hinder effective issue identification and resolution, posing challenges in troubleshooting.	High: Hinders visibility and control	Low: Limited by available tools and access
[44]	Security vulnerabilities: Shared resources increase the potential attack surface, necessitating robust security measures despite inherent vulnerabilities.	High: Exposes systems to potential threats	Low: Can be mitigated with robust security measures
[46]	Performance unpredictability: Variability in execution times and shared resource performance affects application predictability.	High: Hampers application predictability	Low: Variability inherent in shared resources
[20]	Scalability concerns: Limits on scaling IoT workloads due to concurrency and resource allocation constraints pose challenges.	High: Limitations on scaling IoT workloads	Low: Automatic scaling alleviates some constraints
[13]	Cost management complexities: Difficulties in predicting and managing costs arise due to granular billing and resource usage complexities.	High: Difficulties in predicting and managing costs	Low: Granular billing adds complexity
[46]	Integration challenges: Disruptions in interoperability may occur when integrating server-less components with existing systems	High: Disruption in interoperability	Low: Dependent on system architecture
[30]	Lack of control over infrastructure: Reduced control over underlying infrastructure restricts customization and optimization.	High: Restricts customization and optimization	Low: Varied control levels inherent to server-less

[7]	Network limitations: Dependency on provider networks might lead to performance degradation and latency issues	High: Risks performance degradation	Low: Depends on provider network quality
[31]	Limited support for long-running tasks: Constraints on execution duration hinder tasks requiring prolonged computation	High: Constraints on prolonged computations	Low: Can be addressed by task partitioning

The challenges and the severity of their impact in implementing serverless resource management in cloud-based IoT environment are shown in figure 2. According to the analysis, security vulnerabilities, cold start delay, and scalability concerns are some of the most critical factors that impact the system's performance, reliability, and resource utilization.

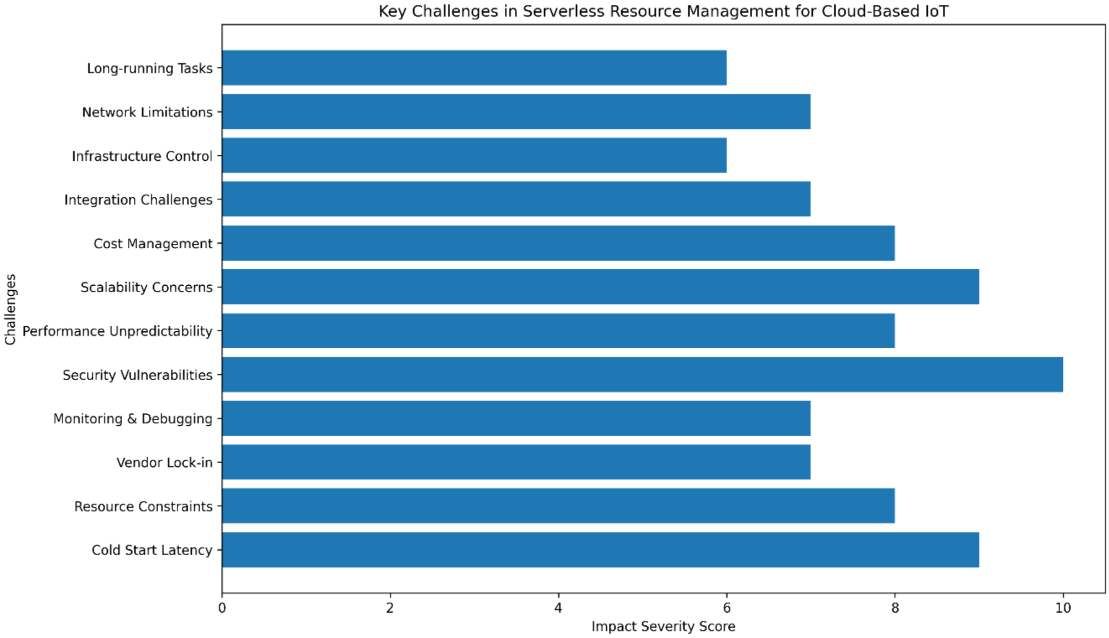

Figure 2. Key challenges in server-less resource management for cloud-base IoT.

The results of this study address the first research question (RQ1) (RQ1) by defining the key challenges and opportunities of serverless based resource management in cloud-based IoT applications. The reviewed studies showed that the issues including cold start latency, resource limitation, vendor lock-in, security vulnerabilities, monitoring and debugging complexity, scalability limitations, performance unpredictability, and others have significant impacts on performance and reliability of IoT-cloud systems. Meanwhile, the study also revealed that serverless computing offers certain opportunities such as better scalability, dynamic resource provisioning, cost-efficient management of infrastructure, lower operational overhead and better load-balancing. In addition, the analysis revealed that there are techniques that can improve system performance, throughput, energy efficiency and resource utilization, such as predictive scaling, hybrid scaling, auto-scaling policies and cost-aware resource provisioning. The presented results collectively address the first research question by showing that, while serverless computing presents a few technical and operational challenges, it has significant potential for efficient, scalable, and cost-effective resource management resources in modern cloud based IoT environments with the help of intelligent scaling and security mechanisms.

RQ2: What role do dynamic scaling strategies play in achieving cost-efficient resource utilization in cloud-based IoT environments, and how can these strategies be optimized for different application scenarios?

Effectively managing resources in cloud-based IoT environments requires a nuanced approach to dynamic scaling strategies. Predictive scaling anticipates future needs but grapples with accuracy in dynamic IoT settings. These auto-scaling policies are adjustable on time, rendering them more detail-oriented than others. There are different strategies that might be helpful like scaling horizontally or vertically, but another discipline that have to confront is the bottleneck or limited scales. Hybrid scaling, in turn, tries to create the balance of these paradigms. Irresponsive scaling, on the other hand, is quick enough, but it may overprovision in case the network is not prepared to accommodate this. Scale implementation based on rules yields structural design, but is not reliable for uncertain situations. Cost-conscious scaling combines resource optimization with financial concerns; specific application will give way to the application of strategies that take into account the most specific profiling. Flexible period operation might not be able to respond to changes effectively in contrast with time-based scaling. These strategies collectively provide us with an enormous toolbox, but the hardest part is the fact that application of certain strategy is a dynamic process, in which we have to think who and how would that strategy best fit with the specific features of IoT.

Predictive Scaling:

Scaling predictively means having a capacity to foresee or project into the future the need for more resources by using historical data, patterns or predictive algorithms. Targets for this action are to avoid underutilization of facilities and to deliver additional resources just before the moment of demand spikes. Nonetheless, promised high accuracy is only achieved with difficulties, in IoT environments which are fast-variating: with overrunning demand and different workloads and patterns.

Auto-scaling Policies:

Auto-scaling dynamics can automatically resize the resources online in real time in accordance with the requirements of the computing system. This brings in the capacity systems to do the scaling decisions about resource usage in a manner that is cost-effective and appropriate to the workload situation. However, the adaptiveness of auto scaling policies relies on fine-adjustment of the threshold to hit the balance between an emphasize on both speed and efficiency.

Horizontal Scaling:

To be more precise, vertical scaling means adding or removing resources, such as memory or computing power, in order to match the changing workload. It exerts the burden of work across various instances so running stronger traffic load is possible. Nevertheless, system bottlenecking within and distributed networks can hamper systems workload, affecting the performance of the system as a whole.

Vertical Scaling:

Vertical scaling expands the effectiveness of odd available resources by increasing the load capacity of single instances. The auto-scaling strategy has a very nice ability to adapt to the workloads in terms of scaling up or down the resources of an individual instance as needed. Yet, its high rate of efficiency can be a drawback because of certain significant capacity limitations related to vertical scaling approaches.

Hybrid Scaling:

Hybrid scaling in this case war is the ability to employ both horizontal and vertical scaling approaches to leverage the advantages while eliminating the disadvantages at the same time. With this capability, it pools compute resources, be it computer instances or computing power, to dynamically add instances and enhance individual instance capacity. The problem that might appear is how to govern the one of the two formations (e.g. the hybrid one) which may be difficult, so it is necessary to properly control and monitor the system.

Reactive Scaling:

In contrast with proactive scaling, the reactive scaling is a scaling method that deal with the immediate fluctuations of workload by very quickly adjusting speedily resources. This, at the same time, can be one of the perfect techniques for absorbing fluctuations in demand but also can give rise to the problem of over-provisioning through the very fluctuations, the efficiency of which could be very weak.

Rule-based Scaling:

The rule-based scaling approach is carried out in accordance with the set rules or triggering of thresholds in the scaling decisions. It generally is a structural framework for nonlinear scaling actions that brings forth amendments for particular conditions. Although the strength is based on a narrow spectrum of circumstances in which exogenous variables are efficiently disposed, this determinism may be problematic and make it impossible to handle dynamic scenarios outside the scope of predefined rules.

Cost-aware Scaling:

Cost-conscious scaling weighs the cost perspectives while deciding how to scale the business units. The idea is to come up with the best solution using available resources at a minimal cost. Nevertheless, the task is the most difficult part because it is unreal to really assess the real-time costs which dynamic and changes in a changing environment all the time [55].

Application-specific Scaling:

Application oriented scaling makes use of all the available resources of the cloud provider and tweaks scaling strategies to the unique requirement in case of specific application. Here high accuracy required to conduct proper profiling of the applicants, as it is about the best GCM to enforce at each area. The hurdle is the extent of the tracking and a coverall determining whether to implement or not the specific of this strategy.

Time-based Scaling:

When the processes are time-based, resources are scaled according to the pre-set time intervals or schedules, saving time and preventing possible mistakes. This approach of planning and then adjusting ensures that the systems are working adequately, but these systems lack flexibility to respond to sudden changes in the workload outside the schedule, which affects responsiveness and adaptability in dynamic restricted environments. Table 6 demonstrated the Strategies and their importance associated with implementing server-less computing for resource management in cloud-based IoT. This table also highlights the limitations associated with each strategy.

Table 6. Strategies and the importance of studies

Research Study	Strategies	Importance	Limitation
[44]	Predictive scaling: Anticipates resource needs based on historical data or predictive algorithms to reduce underutilization, but accuracy hinges on precise predictions in dynamic environments.	High: Foresees resource needs, reducing underutilization	Low: Complexity in accurate predictions
[1]	Auto-scaling policies: Dynamically adjusts resources in real-time based on demand, automating scaling decisions to match workload requirements, albeit requiring fine-tuning for optimal efficiency.	High: Adjusts resources based on real-time demands	Low: Challenges in fine-tuning policies

[12]	Horizontal scaling: Distributes workload across multiple instances to manage varying demands efficiently but may introduce potential bottlenecks in distributed systems.	High: Adds or removes instances based on load	Low: Potential bottlenecks in distributed systems
[5]	Vertical scaling: Enhances individual instance capacity for workload management, but scalability may be limited compared to horizontal scaling due to instance-centric optimization.	High: Optimizes existing resources efficiently	Low: Limited scalability compared to horizontal scaling
[3]	Hybrid scaling: Balances workload demands by combining horizontal and vertical scaling, yet managing hybrid configurations can be intricate due to the blend of strategies.	High: Balances between horizontal and vertical scaling	Low: Complexity in managing hybrid configurations
[4]	Reactive scaling: Swiftly adjusts resources in response to immediate workload changes, although this agility may lead to potential over-provisioning during fluctuations.	High: Rapidly responds to immediate workload changes	Low: Potential over-provisioning during fluctuations
[16]	Rule-based scaling: Makes scaling decisions based on predefined rules or thresholds, providing structure but potentially struggling with unforeseen scenarios outside predefined parameters.	High: Follows predefined rules for scaling decisions	Low: Rigidness in handling dynamic scenarios
[55]	Cost-aware scaling: Considers cost implications in scaling decisions to optimize resource utilization, requiring real-time cost assessment in dynamic environments.	High: Considers cost implications in scaling decisions	Low: Complexity in real-time cost assessment
[13]	Application-specific scaling: Tailors scaling strategies to meet unique application requirements, demanding comprehensive application profiling for precise resource alignment and efficiency.	High: Tailors scaling strategies to application requirements	Low: Requires detailed application profiling
[13]	Time-based scaling: Shifts resource allocation for time intervals or schedules, an automatic action to scale actions in a set period, although the unexpected demand might not be swiftly tracked and the ability to adjust is limited.	High: Adjusts resources based on predefined time intervals	Low: Limited adaptability to sudden workload changes

The comparative performance of different approaches to dynamic scaling for cost-efficient resource usage and performance optimization in cloud-based IoT environments is shown in Figure 3. The analysis also suggests that auto-scaling, predictive scaling, hybrid scaling, and cost-aware scaling are some of the best options for managing dynamic workloads, increasing scalability, and optimizing resource allocation within an IoT-cloud infrastructure.

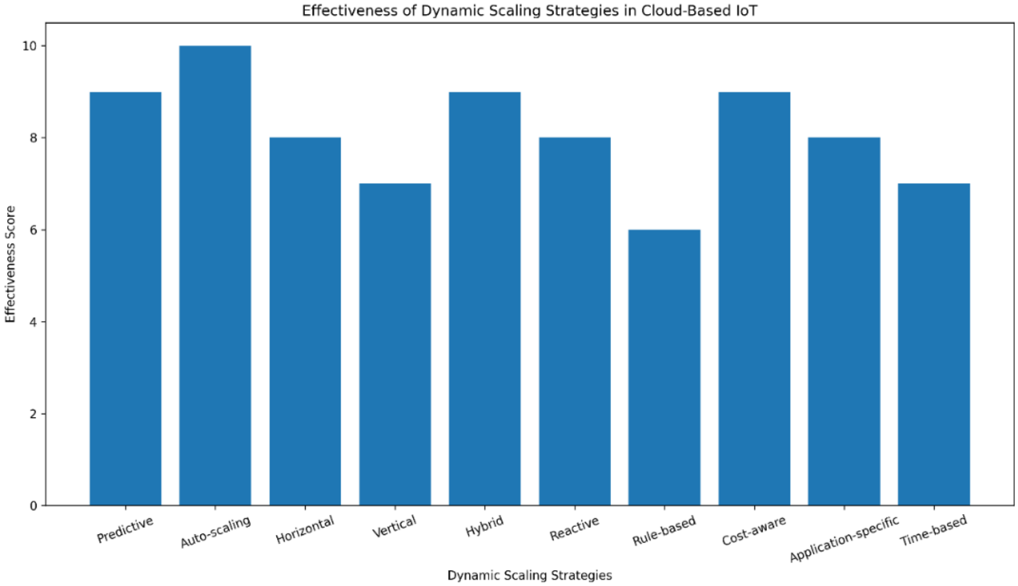

Figure 3. Effectiveness of dynamic scaling strategies in cloud-based IoT.

The results of this study address our second research question (RQ2) and show that dynamic scaling can be essential to use resources efficiently and at scale in cloud-based IoT environments. The reviewed studies showed that several strategies including predictive scaling, auto-scaling policies, horizontal scaling, vertical scaling, hybrid scaling, reactive scaling, rule-based scaling, cost-aware scaling, application-specific scaling, and time-based scaling effectively help to optimize the allocation of resources and workload management in a dynamic IoT-cloud infrastructure. The results also reveal that predictive and auto-scaling strategies contribute to better resource utilization by dynamically scaling resources based on workload predictions and trends, respectively, and that horizontal and vertical scaling are key for scalability and performance. Hybrid scaling was identified to be a balanced option when it comes to combining the benefits of both horizontal and vertical scaling strategies. Additionally, cost-aware scaling and application-specific scaling techniques enable organizations to achieve cost and application-specific optimizations on their resources in accordance with their application needs and workload characteristics. The study identified some of the limitations of these approaches, including inaccurate predictions, pipeline problems in distributed systems, risks of over-provisioning, lack of adaptability, and real-time cost assessment. The overall results answered the second research question and highlighted the importance of dynamic scaling strategies to maximize resource utilization and attain optimal performance in cloud-based IoT systems; however, the choice and adaptation of these strategies in the specific scenario, workload behavior, and infrastructure requirements are important to the success of these strategies.

The taxonomy diagram as shown in Figure 4 outlines the organizational structure of a research study titled "Cost-Effective & Scalable Resource Management in Cloud-Based IoT. “The paper opens with the Introduction, which lays the groundwork through two key elements: Background, which offers historical context and foundational information on cloud-based IoT systems, and Detail, which elaborates on the particular focus and importance of the research. The next major section addresses Issues & Challenges, divided into four main segments: Challenges, which identify technical obstacles and practical complications in achieving efficient and scalable resource use; Resource Management, which reviews current practices and frameworks for controlling and distributing IoT resources; Effectiveness, which examines how well these approaches fulfill objectives related to performance and cost savings; and Assessment, which involves the evaluation, comparison, or analysis of existing techniques and solutions.

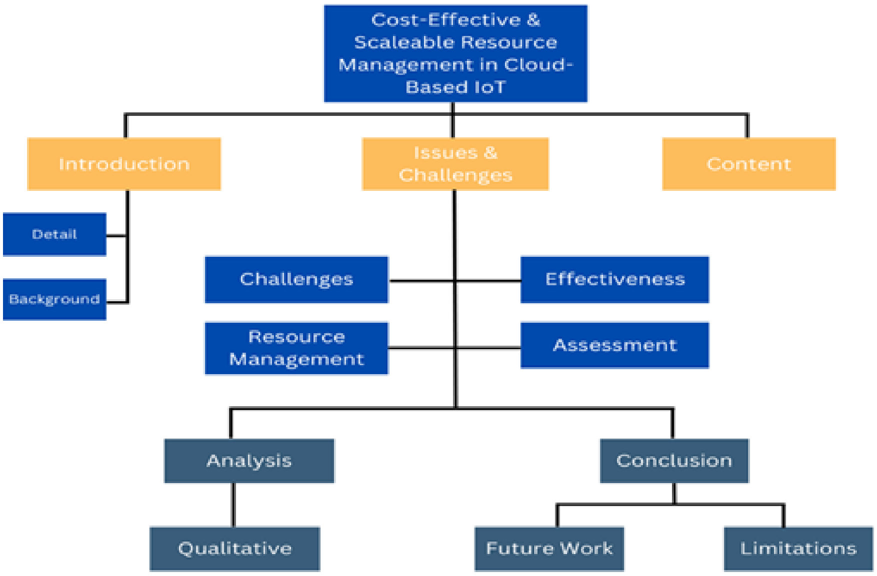

Figure 4. Taxonomy of cost-effective and scalable resource management in cloud-based IoT.

The content section further supports the analytical discussion into the evaluative aspects of the paper. This leads to the Analysis portion, which uses a Qualitative methodology to interpret insights, observations, and patterns based on descriptive, non-statistical data. The study closes with the Conclusion, which consists of two final parts: Future Work, which proposes areas for continued exploration or enhancement, and Limitations, which outline the boundaries, weaknesses, or restrictions of the current investigation. Altogether, this taxonomy presents a well-structured and systematic flow that guides the reader from foundational concepts to key findings and future directions.

Practical Implications:

This research has practical implications for real-life IoT-Cloud applications in different industrial sectors like healthcare, smart cities, manufacturing, transportation, agriculture, and intelligent surveillance systems. In large-scale IoT deployments, the identified resource management and dynamic scaling approaches can assist organizations in efficiently managing their workloads, optimizing computational resources, and minimizing operational costs. Better scalability, lower latency, higher throughput, and reliable services are provided by the techniques of cloud systems to dynamically allocate resources according to real-time demand, such as predictive scaling, auto-scaling, hybrid scaling, and cost-aware resource provisioning. Further, by leveraging the serverless paradigm and the adaptive management of resources, the overhead of infrastructure and the energy consumption in distributed IoT systems can be minimized. Industrially, these insights can assist decision makers and/or system architects in selecting a scaling or resource optimization technique for different requirements, workload characteristics, and constraints. Moreover, the research highlights the importance of using a mixture of security mechanisms, monitoring, and workload balancing technologies to ensure reliable and sustainable IoT-cloud operations. The proposed contributions and comparative analyses provide a viable foundation for the design of intelligent, scalable, and cost-effective IoT-cloud systems to enable modern industrial use cases and future smart environments.

Conclusions:

This paper delivers helpful instructions to establish effective resource management solutions for cloud-based IoT environments. Main findings pointed out resource allocation strategies as lightweight and fast which make it a critical component of the infrastructure of the future Security flew out as a significant element of consideration, which highlighted key management protocols and encryption algorithms related to securing delicate details. The obtained data provides significant implications, which affect future study development and practical system deployment. System developers building IoT implementations should use these techniques to create platforms, which are more effective and protected, as well as flexible systems. Researches should study new technologies like edge computing and federated learning to discover ways, which solve present resource challenges and open future prospects in resource management areas.

Future Directions:

The research areas that can be served in the future include the investigation of edge computing impact on resource management, researching of new encryption techniques to boost data security as well as examination of federated learning scalability approaches in the distributed IoT environments, by tackling the lingering concerns and grabbing the given opportunities, the subject area of cloud-based IoT resource management will continue with its evolution and sticks to the podium of innovation along with developing the capabilities of the connected devices.

Future research in resource management for cloud-based IoT environments should focus on the development of intelligent, adaptive, and secure resource orchestration frameworks capable of handling highly dynamic and heterogeneous IoT workloads. Emerging technologies such as edge intelligence, federated learning, reinforcement learning, and AI-driven predictive scaling should be further explored to improve real-time decision-making, latency reduction, energy efficiency, and autonomous resource allocation. In addition, future studies should investigate lightweight security and privacy-preserving mechanisms for distributed IoT-cloud systems, particularly in healthcare, smart city, and industrial IoT applications where data sensitivity and reliability are critical. Researchers should also emphasize the integration of serverless computing, hybrid fog–cloud architectures, and sustainable green computing models to achieve scalable and cost-aware IoT infrastructures. Furthermore, more empirical evaluations, real-world industrial deployments, and benchmark-driven comparative studies are needed to validate the effectiveness, interoperability, and long-term sustainability of proposed resource management strategies in practical IoT-cloud ecosystems.

References:

[1] Yicheng Yu, Liang Hu, "A Secure Authentication and Key Agreement Scheme for IoT-Based Cloud Computing Environment," *Symmetry (Basel).*, vol. 12, no. 1, p. 150, 2020, doi: https://doi.org/10.3390/sym12010150.

[2] M. El Idrissi, O. Elbeqqali, and J. Riffi, "A Review On Relationship Between Iot-Cloud Computing - Fog Computing (Applications And Challenges)," *2019 3rd Int. Conf. Intell. Comput. Data Sci. ICDS 2019*, Oct. 2019, doi: 10.1109/ICDS47004.2019.8942304.

[3] Jasenka Dizdarević, Francisco Carpio, "A Survey of Communication Protocols for Internet of Things and Related Challenges of Fog and Cloud Computing Integration," *ACM Comput. Surv.*, vol. 51, no. 6, pp. 1–29, 2019, [Online]. Available: https://dl.acm.org/doi/10.1145/3292674

[4] L. Minh Dang, Md Jalil Piran, "A Survey on Internet of Things and Cloud Computing for Healthcare," *Electronics*, vol. 8, no. 7, p. 768, 2019, doi: https://doi.org/10.3390/electronics8070768.

[5] Waleed Noori Hussein, Haider N. Hussain, "A proposed framework for healthcare based on cloud computing and IoT applications," *Mater. Today Proc.*, 2022, doi: 10.1016/j.matpr.2021.12.505.

[6] "Resource Management in Cloud and Cloud-influenced Technologies for Internet of Things Applications | ACM Computing Surveys." Accessed: May 13, 2026. [Online]. Available: https://dl.acm.org/doi/10.1145/3571729

[7] Lei Li, Mian Guo, "Online Workload Allocation via Fog-Fog-Cloud Cooperation to Reduce IoT Task Service Delay," *Sensors*, vol. 19, no. 18, p. 3830, 2019, doi: https://doi.org/10.3390/s19183830.

[8] M. Aleisa, A. A. Hussein, F. Alsubaei, and F. T. Sheldon, "Performance Analysis of Two Cloud-Based IoT Implementations: Empirical Study," *Proc. - 2020 7th IEEE Int. Conf. Cyber Secur. Cloud Comput. 2020 6th IEEE Int. Conf. Edge Comput. Scalable Cloud, CSCloud-EdgeCom 2020*, pp. 276–280, Aug. 2020, doi: 10.1109/CSCLOUD-EDGECOM49738.2020.00055.

[9] "A based cloud computing solution using DiffServ architecture for scalability issue in IOT networks with multiple SLA requirements - Détails de la communication." Accessed: May 13, 2026. [Online]. Available: https://ucarech.uca.ma/open-research/communication.php?id=11164

[10] V. N. Pham and E. N. Huh, "An efficient edge-cloud publish/subscribe model for large-scale IoT applications," *Adv. Intell. Syst. Comput.*, vol. 935, pp. 130–140, 2019, doi: 10.1007/978-3-030-19063-7_12/SAVE-RESEARCH.

[11] Mohamed Amine Ghamri, Badis Djamaa, "Reinforcement Learning based collaborative DNN inference for edge intelligence," *Futur. Gener. Comput. Syst.*, vol. 174, p. 107942, 2026, doi: https://doi.org/10.1016/j.future.2025.107942.

[12] S. K. and V. P. S. MANDEEP KAUR, "Critical review of security issues of Internet of Things under cloud computing environment," *Adv. Appl. Math. Sci.*, vol. 18, no. 8, pp. 627–636, 2019, [Online]. Available: https://www.mililink.com/upload/article/1067210990aams_vol188_june_2019_a5_p627-636_mandeep_kaur,_sanchi_kakkar_and_v.p._singh.pdf

[13] S. -Y. Kim and H. Ko, "Distributed Split Computing System in Cooperative Internet of Things (IoT)," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3298810.

[14] I. Wang, E. Liri, and K. K. Ramakrishnan, "Supporting IoT Applications with Serverless Edge Clouds," *Proc. - 2020 IEEE 9th Int. Conf. Cloud Networking, CloudNet 2020*, Nov. 2020, doi: 10.1109/CLOUDNET51028.2020.9335805.

[15] R. Sikarwar, P. Yadav, and A. Dubey, "A survey on IOT enabled cloud platforms," *Proc. - 2020 IEEE 9th Int. Conf. Commun. Syst. Netw. Technol. CSNT 2020*, pp. 120–124, Apr. 2020, doi: 10.1109/CSNT48778.2020.9115735.

[16] "(PDF) An Edge-Fog-Cloud Architecture of Streaming Analytics for Internet of Things Applications." Accessed: May 13, 2026. [Online]. Available: https://www.researchgate.net/publication/335235978_An_Edge-Fog-Cloud_Architecture_of_Streaming_Analytics_for_Internet_of_Things_Applications

[17] M. Dai, Z. Su, R. Li, and S. Yu, "A Software-Defined-Networking-Enabled Approach for Edge-Cloud Computing in the Internet of Things," *IEEE Netw.*, vol. 35, no. 5, pp. 66–73, Sep. 2021, doi: 10.1109/MNET.101.2100052.

[18] P. Velmurugadass, S. Dhanasekaran, and S. Sasikala, "The Cloud based Edge Computing with IoT Infrastructure and Security," *2021 Int. Conf. Comput. Perform. Eval. ComPE 2021*, pp. 30–34, 2021, doi: 10.1109/COMPE53109.2021.9751942.

[19] S. S. Salunkhe *et al.*, "An incremental learning on cloud computed decentralised IoT devices," *Int. J. Eng. Syst. Model. Simul.*, vol. 14, no. 1, pp. 1–7, 2023, doi: 10.1504/IJESMS.2023.127397.

[20] "Role of IoT-Cloud Ecosystem in Smart Cities : Review and Challenges - ScienceDirect." Accessed: May 13, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214785320376215

[21] J. Fang and A. Ma, "IoT Application Modules Placement and Dynamic Task Processing in Edge-Cloud Computing," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12771–12781, Aug. 2021, doi: 10.1109/JIOT.2020.3007751.

[22] M. Baqer, "Energy-Efficient Federated Learning for Internet of Things: Leveraging In-Network Processing and Hierarchical Clustering," *Futur. Internet*, vol. 17, no. 1, p. 4, 2025, doi: https://doi.org/10.3390/fi17010004.

[23] "(PDF) Artificial Intelligence in SDN-Enabled Internet of Vehicles: A Comprehensive Survey." Accessed: May 19, 2026. [Online]. Available: https://www.researchgate.net/publication/399183847_Artificial_Intelligence_in_SDN-Enabled_Internet_of_Vehicles_A_Comprehensive_Survey

[24] P. Karthikeyan and K. Brindha, "Smart data flow: a trust-driven hybrid fragmentation framework for optimizing data transmission in IoT-enabled edge-fog-cloud systems," *Computing*, vol. 107, no. 6, Jun. 2025, doi: 10.1007/s00607-025-01491-2.

[25] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[26] X. Zheng, S. Xue, H. Cao, F. Wang, and M. Zhang, "A Cost-efficient Smart IoT Device Controlling System Based on Bluetooth Mesh and Cloud Computing," *Proc. - 2020 Chinese Autom. Congr. CAC 2020*, pp. 3374–3379, Nov. 2020, doi: 10.1109/CAC51589.2020.9326634.

[27] S. P. Ahuja and N. Wheeler, "Architecture of Fog-Enabled and Cloud-Enhanced Internet of Things Applications," *Int. J. Cloud Appl. Comput.*, vol. 10, no. 1, 2020, doi: 10.4018/IJCAC.2020010101.

[28] A. Amairah, B. N. Al-Tamimi, M. Anbar, and K. Aloufi, "Cloud computing and internet of things integration systems: A review," *Adv. Intell. Syst. Comput.*, vol. 843, pp. 406–414, 2019, doi: 10.1007/978-3-319-99007-1_39/SAVE-RESEARCH.

[29] F. Alhaidari, A. Rahman, and R. Zagrouba, "Cloud of Things: architecture, applications and challenges," *J. Ambient Intell. Humaniz. Comput. 2020 145*, vol. 14, no. 5, pp. 5957–5975, Aug. 2020, doi: 10.1007/S12652-020-02448-3.

[30] S. Ketu and P. K. Mishra, "Cloud, Fog and Mist Computing in IoT: An Indication of Emerging Opportunities," *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 39, no. 3, pp. 713–724, 2022, doi: 10.1080/02564602.2021.1898482.

[31] "(PDF) Cyber-Storms Come from Clouds: Security of Cloud Computing in the IoT Era." Accessed: May 13, 2026. [Online]. Available: https://www.researchgate.net/publication/333603006_Cyber-Storms_Come_from_Clouds_Security_of_Cloud_Computing_in_the_IoT_Era

[32] M. De Donno, K. Tange and N. Dragoni, M. De Donno, K. Tange and N. Dragoni, "Foundations and Evolution of Modern Computing Paradigms: Cloud, IoT, Edge, and Fog," *IEEE Access*, vol. 7, pp. 150936–150948, 2019, doi: 10.1109/ACCESS.2019.2947652.

[33] C. Zhu, V. C. M. Leung, L. Shu, and E. C. H. Ngai, "Green Internet of Things for Smart World," *IEEE Access*, vol. 3, pp. 2151–2162, 2015, doi: 10.1109/ACCESS.2015.2497312.

[34] M. Cui, S. S. Baek, R. G. Crespo, and R. Premalatha, "Internet of things-based cloud computing platform for analyzing the physical health condition," *Technol. Health Care*, vol. 29, no. 6, pp. 1233–1247, 2021, doi: 10.3233/THC-213003.

[35] M. Merzouki, C. Mahmoudi, R. Bohn, and C. Tunc, "Security automation for cloud-based iot platforms," *Proc. - 2019 IEEE Int. Congr. Cybermatics 12th IEEE Int. Conf. Internet Things, 15th IEEE Int. Conf. Green Comput. Commun. 12th IEEE Int. Conf. Cyber, Phys. S…*, pp. 1185–1191, Jul. 2019, doi: 10.1109/ITHINGS/GREENCOM/CPSCOM/SMARTDATA.2019.00199.

[36] "(PDF) Investigating How the Cloud Computing Transforms the Development of Industries." Accessed: May 13, 2026. [Online]. Available: https://www.researchgate.net/publication/345380735_Investigating_How_the_Cloud_Computing_Transforms_the_Development_of_Industries

[37] "(PDF) Transformative Effects of IoT, Blockchain and Artificial Intelligence on Cloud Computing: Evolution, Vision, Trends and Open Challenges." Accessed: May 13, 2026. [Online]. Available: https://www.researchgate.net/publication/335938628_Transformative_Effects_of_IoT_Blockchain_and_Artificial_Intelligence_on_Cloud_Computing_Evolution_Vision_Trends_and_Open_Challenges

[38] "Lightweight IoT-based authentication scheme in cloud computing circumstance | Request PDF." Accessed: May 13, 2026. [Online]. Available: https://www.researchgate.net/publication/327411930_Lightweight_IoT-based_authentication_scheme_in_cloud_computing_circumstance

[39] N. Almolhis, A. M. Alashjaee, S. Duraibi, F. Alqahtani, and A. N. Moussa, "The Security Issues in IoT-Cloud: A Review," *Proc. - 2020 16th IEEE Int. Colloq. Signal Process. its Appl. CSPA 2020*, pp. 191–196, Feb. 2020, doi: 10.1109/CSPA48992.2020.9068693.

[40] R. Munoz *et al.*, "Orchestration of Optical Networks and Cloud/Edge Computing for IoT Services," *OECC/PSC 2019 - 24th Optoelectron. Commun. Conf. Conf. Photonics Switch. Comput. 2019*, Jul. 2019, doi: 10.23919/PS.2019.8817786.

[41] Y. Fan, G. Zhao, W. Shang, J. Shang, W. Lin, and Z. Wang, "A Preliminary Design for Authenticity of IoT Big Data in Cloud Computing," *Proc. - Int. Conf. Comput. Commun. Networks, ICCCN*, vol. 2020-August, Aug. 2020, doi: 10.1109/ICCCN49398.2020.9209646.

[42] I. Ahammad, A. R. Khan, and Z. Us Salehin, "A review on cloud, fog, roof, and dew computing: IoT perspective," *Int. J. Cloud Appl. Comput.*, vol. 11, no. 4, pp. 14–41, Oct. 2021, doi: 10.4018/IJCAC.2021100102.

[43] M. S. Jassas and Q. H. Mahmoud, "Evaluation of Failure Analysis of IoT Applications Using Edge-Cloud Architecture," *SysCon 2022 - 16th Annu. IEEE Int. Syst. Conf. Proc.*, 2022, doi: 10.1109/SYSCON53536.2022.9773898.

[44] A. Alnoman, S. K. Sharma, W. Ejaz, and A. Anpalagan, "Emerging edge computing technologies for distributed IoT systems," *IEEE Netw.*, vol. 33, no. 6, pp. 140–147, Nov. 2019, doi: 10.1109/MNET.2019.1800543.

[45] M. N. Bahiri, A. Zyane, and A. Ghammaz, "An enhancement for the autonomic middleware-level scalability management within IoT system using cloud computing," *Lect. Notes Electr. Eng.*, vol. 519, pp. 80–88, 2019, doi: 10.1007/978-981-13-1405-6_11/SAVE-RESEARCH.

[46] "The importance of integrating Internet of Things, big data and cloud computing into linguistic landscapes – JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES." Accessed: May 13, 2026. [Online]. Available: https://www.journalimcms.org/special_issue/the-importance-of-integrating-internet-of-things-big-data-and-cloud-computing-into-linguistic-landscapes/

[47] Y. Kotb, I. Al Ridhawi, M. Aloqaily, T. Baker, Y. Jararweh, and H. Tawfik, "Cloud-Based Multi-Agent Cooperation for IoT Devices Using Workflow-Nets," *J. Grid*

Comput., vol. 17, no. 4, pp. 625–650, Dec. 2019, doi: 10.1007/S10723-019-09485-Z/METRICS.

[48] H. Yang and Y. Li, "RETRACTED ARTICLE: Exploration of police vocational training mode based on face recognition technology in the context of IoT cloud computing(Soft Computing, (2024), 28)," *Soft Comput.*, vol. 28, no. Suppl 2, p. 585, Dec. 2024, doi: 10.1007/S00500-023-08436-X;SUBPAGE:STRING:BASIC.

[49] Mohammed Al Masarweh, Tariq Alwada'n, "Fog Computing, Cloud Computing and IoT Environment: Advanced Broker Management System," *J. Sens. Actuator Netw*, vol. 11, no. 4, p. 84, 2022, doi: https://doi.org/10.3390/jsan11040084.

[50] N. Almurisi and S. Tadisetty, "Cloud-based virtualization environment for IoT-based WSN: solutions, approaches and challenges," *J. Ambient Intell. Humaniz. Comput. 2022 1310*, vol. 13, no. 10, pp. 4681–4703, Mar. 2022, doi: 10.1007/S12652-021-03515-Z.

[51] "(PDF) Security Issues in IoT and Cloud Computing Service Models with Suggested Solutions." Accessed: May 13, 2026. [Online]. Available: https://www.researchgate.net/publication/359925487_Security_Issues_in_IoT_and_Cloud_Computing_Service_Models_with_Suggested_Solutions

[52] R. Mafamane, M. Ouadou, A. T. J. Hassani, and K. Minaoui, "Study of the heterogeneity problem in the Internet of Things and Cloud Computing integration," *2020 10th Int. Symp. Signal, Image, Video Commun. ISIVC 2020*, Apr. 2021, doi: 10.1109/ISIVC49222.2021.9487539.

[53] Mir Salim Ul Islam, Ashok Kumar, "Context-aware scheduling in Fog computing: A survey, taxonomy, challenges and future directions," *J. Netw. Comput. Appl.*, vol. 180, p. 103008, 2021, doi: https://doi.org/10.1016/j.jnca.2021.103008.

[54] "(PDF) Latency-Aware Task Scheduling for IoT Applications Based on Artificial Intelligence with Partitioning in Small-Scale Fog Computing Environments." Accessed: May 19, 2026. [Online]. Available: https://www.researchgate.net/publication/363945254_Latency-Aware_Task_Scheduling_for_IoT_Applications_Based_on_Artificial_Intelligence_with_Partitioning_in_Small-Scale_Fog_Computing_Environments

[55] I. Mohiuddin and A. Almogren, "Security Challenges and Strategies for the IoT in Cloud Computing," *2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020*, pp. 367–372, Apr. 2020, doi: 10.1109/ICICS49469.2020.239563.