Journal of Information Systems and Informatics
Vol. 6, No. 1, March 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i1.640 Published By DRPM-UBD

Data Analytics Techniques for Addressing Cloud Computing Resources Allocation Challenges: A Bibliometric Analysis Approach

Sello Prince Sekwatlakwatla[1], Vusumuzi Malele[2]

[1,2]North-West University, Potchefstroom, South Africa
Email: sek.prince@gmail.com[1], vusi.Malele@nwu.ac.za[2]

Abstract

The increase in the use of digital technology led to an increase in online activities. In this regard, many organizations adopted cloud computing systems to manage this online traffic. It is plan of every cloud computing resource provider to manage their system effectively and efficiently. This paper uses bibliometric analysis technique to look at the prevalence of utilization of data analytics techniques in addressing cloud computing resource allocation challenges. In this regard, the following research databases the Association for Computing Machinery, the Institute of Electrical and Electronics Engineering, Web of Science and Scopus databases, were consulted. The research articles published before the beginning of 2017 to 2023 were considered as part of the analysis. The results showed that the prevalent data analytics techniques used to address the cloud computing resources allocation challenge are Support Vector Machine, Spatio-temporal and edge-cloud collaborative scheme. Failure to effectively and efficiently provide cloud computing management resource allocation will lead to system bottlenecks especially during peak periods. In this regard, such a failure could lead to dissatisfied clients.

Keywords: Cloud computing, Traffic flow, Resources allocation, Data Analytics

1. INTRODUCTION

Cloud computing is a service-oriented technology that provides infrastructure, platforms, and software as a service, with resource allocation based on computer pooling, Stochastic Dynamic Scheduling in Edge-Cloud Environments utilizes Residual Recurrent Neural Networks (R2N2) frameworks to decrease energy, response time, and cost, but does not scale [1-3].

The global adoption of cloud technology has surged in recent years, with a growing number of small and medium-sized businesses realizing its benefits [4]. Companies like Google, Facebook, Amazon, Adobe, Netflix, and many more, are leveraging these benefits, which include cost efficiency, reduced downtime, environmental benefits, and easier collaboration. For example, cloud computing was used for managing and analyzing industrial IoT data in the intelligent digital twin system of which its performance convinces them that the resource is cost effective [5].

Furthermore, a new cloud computing task scheduling method was developed as a resource allocation algorithm utilizing an improved genetic firework algorithm was proposed in [6-7]. Its application led to demonstration of efficient computing time reduction.

Against this backdrop, cloud computing providers saw an increase in cloud computing traffic and this increase creates a challenge of lack of effective and efficient cloud computing resource allocation [8]. The latter challenge needs a solution(s) because an increase in cloud computing traffic causes the system to be slow and might create lack of client's trust towards providers [9].

In recent years, the usage of data analytics techniques was utilized to understand the past (i.e. what happened and why it happened) and predict the future (i.e. what will happen and what should be done) of cloud computing resource allocation [10-11]. For example, a reinforcement learning algorithm for efficient virtual machine allocation on cloud computing systems that utilize multi-servers was proposed [12]. However, it faced a scalability limitation due to the exponential increase cloud traffic. In this regard, a hierarchical solution for virtual machine resource allocation, focusing on minimizing latency degradation and optimizing power consumption was proposed [13]. The work aimed to efficiently allocate virtual machines and provide a solution for data placement in cloud environments [14]. The solution includes a global tier using deep reinforcement learning and a local tier using long short-term memory (LSTM) network [15-16].

Current Data Analytics techniques, such as time series and regression analysis, have limitations in accurately addressing resources allocation challenges due to their assumptions about future traffic patterns and the difficulty in selecting appropriate independent variables for complex network environments [17]. It is becoming increasingly difficult to allocate cloud computing resources, and this is affecting a large proportion of organizations. Small and medium enterprises (SMEs) are most likely to be affected by these challenges [18-19].

In this paper, available data analytics techniques that address the cloud computing resource allocation challenge are examined using a bibliometric analysis technique as discussed in following section. Despite this introduction the article proceeds to section 2 which highlights the adopted methods, then section 3 which provides the study findings and discussion. The conclusion is provided in section IV.

2. METHODS

This paper adopted a qualitative method by collecting secondary data emanating from different research databases. At least four databases were consulted, namely: (i) Association for Computing Machinery (ACM) Digital Library, (ii) Institute of Electrical and Electronics Engineering (IEEE) Xplore, (iii) Web of Service, and

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

(iv) Scopus. The databases were searched for research that relates the solutions that use Data Analytics techniques for managing and addressing cloud computing resource allocation challenges. In this regard, only the publications from five years horizon (i.e., 2017 to 2023) were considered. Otherwise, all other published work from 2016 downwards was not considered.

This search included conference papers, journal articles, and magazines. The following keywords were used: data analytics, cloud analytics, cloud computing, machine learning, and algorithms. Only related papers are included. In this regard, 130 documents with 638 authors were downloaded (see Fig. 1). The Bibliometric analysis technique particularly the thematic maps were used to study the relationship between data analytics techniques and cloud computing resource allocation.

Figure 1. Summary of Analysis

Figure 3 indicates data analytics techniques for addressing cloud computing resource allocation challenges, with cloud computing at 6% and resource allocation at 6%, deep learning at 3%, and performance at 4%.

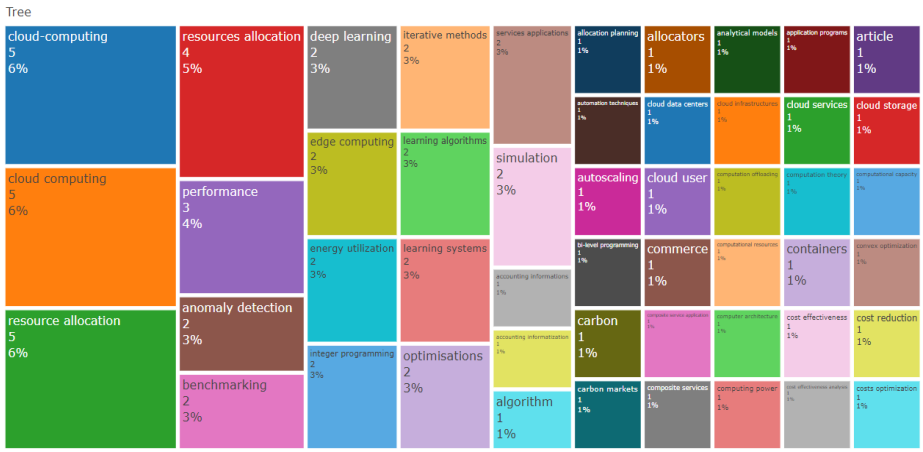

Figure 2. Summary of analysis of cloud computing resources Allocation

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3. RESULTS AND DISCUSSION

Figure 3 indicates the thematic analysis of data analytics techniques for addressing cloud computing resource allocation challenges to organize and analyze complex data sets, focusing on themes that capture the narratives within the system analysis.

3.1. Scopus

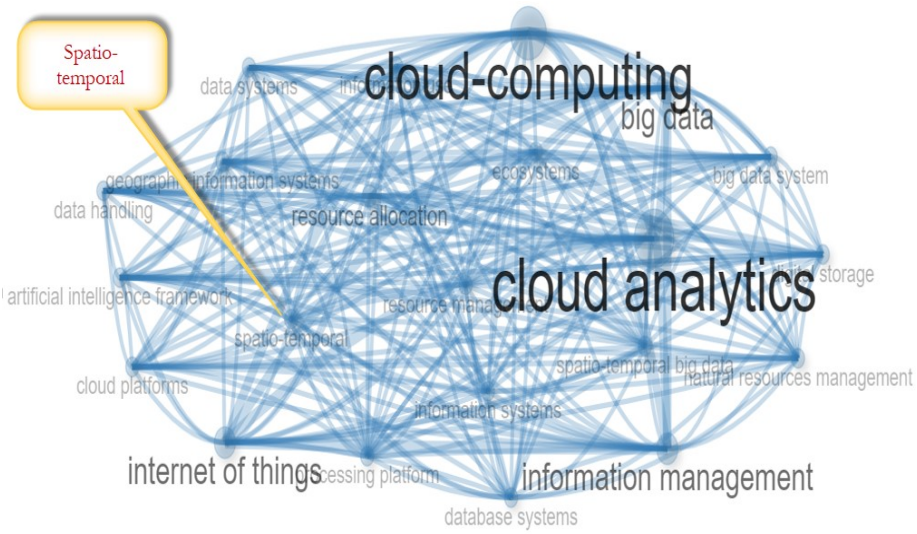

Figure 3. Thematic maps for Scopus database

Thematic results from the bibliometric analysis show data analytics techniques that address computing resources in the cloud, cloud analytics, and cloud computing as main tools that are linked to multiple tools, such as spatio-temporal for big data (see Figure 3).

internet of things and cloud computing, information management, and cloud computing They are playing a critical role in the data analytics techniques for addressing cloud computing resource allocation challenges. However, there was no indication of short-term allocation of resources in cloud computing. Most of these tools are performing very well in traffic prediction, but they have limitations for frequently allocating resources in cloud computing [19-20].

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.2. IEEEXplore

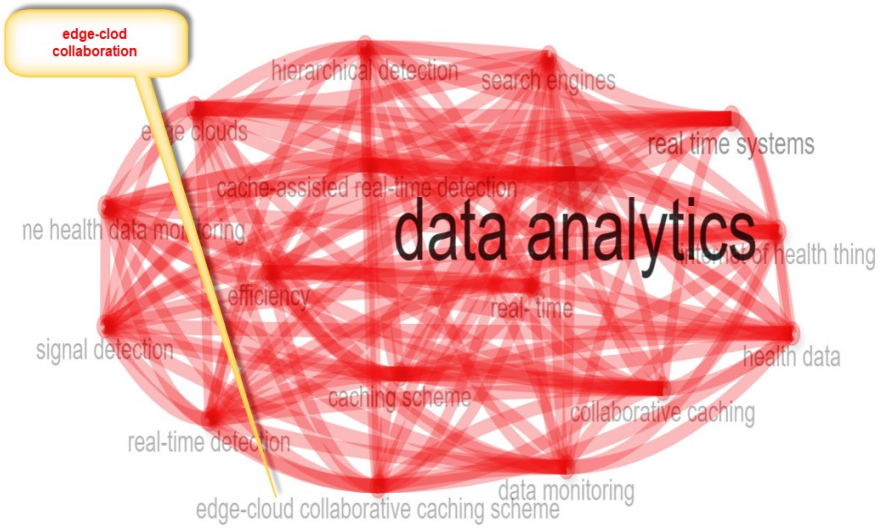

Figure 3. Thematic maps for IEEE database

Figure 3 indicates the thematic results from the IEEE database, data analytics from this database is link to multiple tools to address cloud computing resources allocation concerns, the advantages of the tools highlighted in this database is real-time system, real time detection, signal detection to be use to highlight when the traffic increase, data monitoring tool to increase accuracy of the allocation of the resources, The edge-cloud collaborative caching scheme is proposed as an allocation tool strategy to allocate the resources in real time, and the strategy proved to be effective. However, there is a minimum to no requirement regarding traffic prediction. Future includes a strategy for the privacy of users and the system [13].

3.3. Web of Service

Figure 4 indicates the thematic results from the Web of Service database. In these results, the algorithm is at the center to address computing resources in the cloud, which is supported by optimization. Framework support vector machines is also highlighted as a detection tool in web applications and networks, and at the same time, data analysis architecture is closely linked to performance.

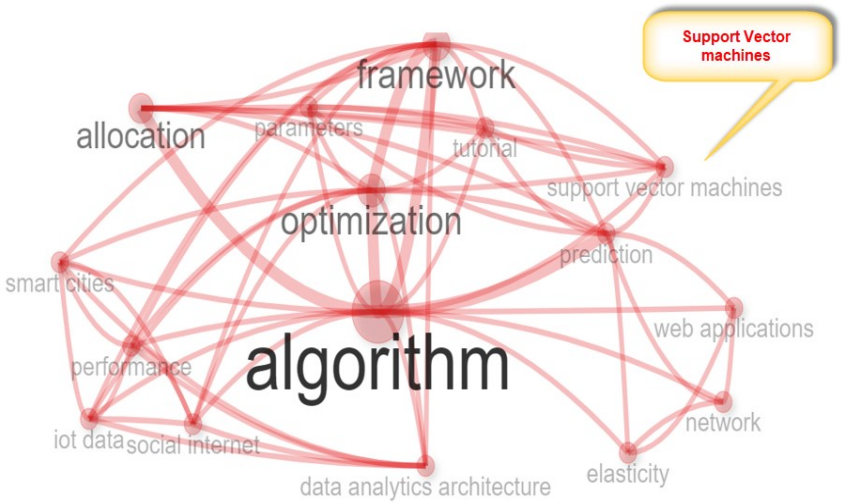

Figure 4. Thematic maps for Web of Service database

Support vector machines enhance resource allocation accuracy and scheduling techniques, achieving fair share of users' essential resources, fulfilling current and future resource allocation requirements in short term traffic flow [14].

3.4. ACM

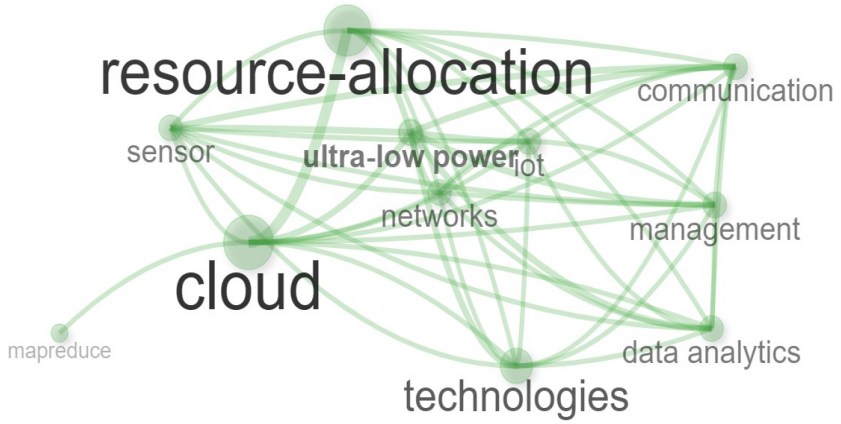

Figure 5. Thematic maps for ACM database

Figure 5 indicates the thematic results from the ACM database. Focusing on the highlights of data analytics techniques that address cloud computing resources allocation in cloud computing, in this result, resource allocation using cloud technology is highlighted and managed by tools such as sensors, ultra-low power,

networks, and communication. In this result, there is a limitation to traffic prediction tools.

3.5. Discussion

Based on the evaluation results of the bibliometric analysis, there are numerous recommendations to enhance cloud computing resource allocation challenges. The results in Section 3.1 show that computing resources in the cloud, cloud analytics, and cloud computing are linked to technique, which is spatio-temporal for big data evaluations, and Section 3.2 results propose the edge-cloud collaborative caching scheme with the advantages of processing data in a real-time system, real-time detection, and signal detection to be used for data monitoring to increase the accuracy of the allocation of resources.

Finally, sections 3.3 and 3.4 propose support vector machines, which assist in cloud computing resource allocation by increasing the performance of the system and enhancing communication strategies. In this regard, this study proposes spatio-temporal cloud collaborative vector machines as a data analytics technique for cloud computing resource allocation (see figure 6).

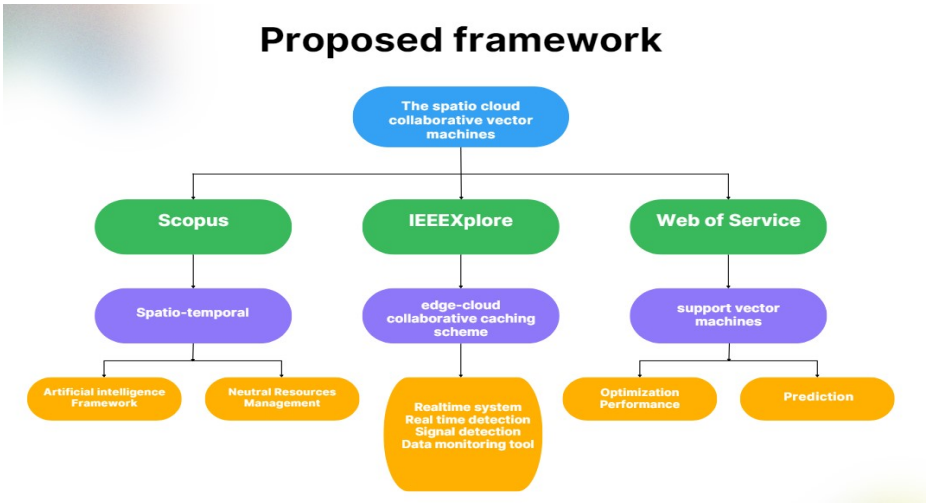

Figure 6. The spatio cloud collaborative vector machines

Figure 6 represents a proposed data analytics technique by combining three techniques that are identified, the spatio cloud collaborative vector machines then merge multiple techniques in order to produce better results. The purpose of combining these techniques is to identify a single model for cloud resource allocation rather than creating one model and hoping that this model will be the best/most precise model for resouce allocation in cloud computing [20].

4. CONCLUSION

Many organizations adopted cloud computing to manage the increase in online activities, which create cloud computing traffic. Unfortunately, an increase in cloud computing traffic creates a challenge for effective and efficient cloud computing resource allocation. This study consulted four databases: the Association for Computing Machinery, the Institute of Electrical and Electronics Engineering, the Web of Service, and Scopus. As the form of inclusion and exclusion criteria, only research papers published from 2017 to 2023 were considered, and those from 2016 were not considered. The review shows that cloud computing providers struggle to allocate resources to meet the demands of their clients, causing bottlenecks, especially during peak periods. The study identified and combined three techniques and proposed spatio cloud collaborative vector machines technique. These methods will ensure that resources are allocated using the advantages of each method. Optimization, performance, and prediction will be supported by support vector machines, and real-time system detection and signaling will be supported by edge cloud collaboration by using enhanced technology. The future study will investigate system security, enhance traffic prediction, and allocate resources.

REFERENCES

[1] E. Marinelli, Y. Yan, V. Magnone, C. Dumargne, P. Barbry, T. Heinis and R. Appuswamy, "Towards Migration-Free "Just-in-Case" Data Archival for Future Cloud Data Lakes Using Synthetic DNA "Proceedings *of the VLDB Endowment*, Vol. 16, pp. 1923–1929, 2023. doi: https://doi.org/10.14778/3594512.3594522

[2] J. McLeod, M. Shepherd, and M. Appelbaum "Evidence of cloud and rainfall modification in a mid-sized urban area – A climatological analysis of Augusta, Georgia", *Journal of City and Environment Interactions*, Vol. 21, 2024. doi: https://doi.org/10.1016/j.cacint.2024.100141

[3] S. Tuli, S. Ilager, K. Ramamohanarao and R. Buyya, "Dynamic Scheduling for Stochastic Edge-Cloud Computing Environments Using A3C Learning and Residual Recurrent Neural Networks," *in IEEE Transactions on Mobile Computing*, vol. 21, no. 3, pp. 940-954, 2022, doi: 10.1109/TMC.2020.3017079.

[4] M. F. Manzoor1, A. Abid, M. S. Farooq, N.A. Nawaz, U. Farooq. "Resource Allocation Techniques in Cloud Computing: A Review and Future Directions", *system engineering, computer technology*, vol. 26, 2020, doi: https://doi.org/10.5755/j01.eie.26.6.25865

[5] K. Raghavendar, I. Batra, A. Malik"A robust resource allocation model for optimizing data skew and consumption rate in cloud based IoT environments", *Decision Analytics Journal,* vol. 7, 2023, doi: https://doi.org/10.1016/j.dajour.2023.100200

[6] S. Wang, T. Zhao, and S. Pang, "Task Scheduling Algorithm Based on Improved Firework Algorithm in Fog Computing," *in IEEE* Access, vol. 8, pp. 32385-32394, 2020, doi: 10.1109/ACCESS.2020.2973758

[7] J. Sheng, Y. Hu, W. Zhou, L. Zhu, B. Jin, J. Wang and X. Wang,"Learning to schedule multi-NUMA virtual machines via reinforcement learning" *Journal of Pattern Recognition,* Vol. 121, 2022, doi: https://doi.org/10.1016/j.patcog.2021.108254

[8] S. Prathiba, S. Sankar" Energy-efficient resource allocation in cloud infrastructure using L3F-MGA and E-ANFIS" *journal of Measurement: Sensors,* Vol. 31, 2024. doi: https://doi.org/10.1016/j.measen.2023.100965

[9] Z. Jin, J. Qian, Z. Kong, C. Pan,"A mobility aware network traffic prediction model based on dynamic graph attention spatio-temporal network", *Journal of Computer Networks,* vol. 235, 2023. doi: https://doi.org/10.1016/j.comnet.2023.109981

[10] Z. Yang, W. Ji, Q. Guo and Z. Wang, "JAVP: Joint-Aware Video Processing with Edge-Cloud Collaboration for DNN Inference"*MM '23: Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp 9152–9160, doi: https://doi.org/10.1145/3581783.3613914

[11] M. Junaid, A. Sohail, F. A. Turjman and R. Ali,"Agile Support Vector Machine for Energy-efficient Resource Allocation in IoT-oriented Cloud using PSO", *ACM Transactions on Internet Technology (TOIT),* Vol. 22, 2021, pp 1–35, doi: https://doi.org/10.1145/3433541

[12] T. Welsh and E. Benkhelifa."On Resilience in Cloud Computing: A Survey of Techniques across the Cloud Domain", *ACM Computing Surveys (CSUR),* Vol. 53, 2020, pp 1–36, doi: https://doi.org/10.1145/3388922

[13] C. Chen, L. Liu, S. Wan, X. Hui and Q. Pei," Data Dissemination for Industry 4.0 Applications in Internet of Vehicles Based on Short-term Traffic Prediction" *ACM Transactions on Internet Technology*, Vol. 22, 2021, pp 1–18, doi: https://doi.org/10.1145/3430505

[14] S. Lin, W. Yang, Y. Hu, Q. Cai, M. Dai, H. Wang, and K. Li,"HPS Cholesky: Hierarchical Parallelized Supernodal Cholesky with Adaptive Parameters" *ACM Transactions on Parallel Computing,* 2023, doi: https://doi.org/10.1145/3630051

[15] V. Cortellessa and L. Traini," Detecting Latency Degradation Patterns in Service-based Systems"ICPE '*Proceedings of the ACM/SPEC International Conference on Performance Engineering*, vol. 20, 2020, pp 161–172, doi: https://doi.org/10.1145/3358960.3379126

[16] H. S. Xie and W. Wang, "Long Short-term Dynamic Graph Neural Networks: for short-term intense rainfall forecasting"MLNLP, Proceedings *of the 5th International Conference on Machine Learning and Natural Language Processing*, vol. 22, 2022, pp 74–80, doi: https://doi.org/10.1145/3578741.3578757

[17] M. Feng, J. Zheng, J. Ren and Y. Liu, "Towards Big Data Analytics and Mining for UK Traffic Accident Analysis, Visualization & Prediction"

ICMLC ', Proceedings of the 12th International Conference on Machine Learning and Computing, vol. 20, 2020, pp 225–229, doi: https://doi.org/10.1145/3383972.3384034

[18] R. T. Elmaghraby, N. M.A. Aziem, M. A. Sobh, A. M. Bahaa-Eldin,"Encrypted network traffic classification based on machine learning"*Ain Shams Engineering Journal,* vol. 15, 2024. doi: https://doi.org/10.1016/j.asej.2023.102361

[19] K. Saidi, O. Hioual and A. Siam, "Resources Allocation in Cloud Computing: A Survey". *In: Hatti, M. (eds) Smart Energy Empowerment in Smart and Resilient Cities. ICAIRES. Lecture Notes in Networks and Systems,* vol. 102, 2020, Springer, Cham. doi: https://doi.org/10.1007/978-3-030-37207-1_37

[20] J. Chen, T. Du and G. Xiao, "A multi-objective optimization for resource allocation of emergent demands in cloud computing", *Journal of Cloud Computing*, Vol.20, 2021. doi: https://doi.org/10.1186/s13677-021-00237-7