

DAFTAR PUSTAKA

- [1] P. M. Mell and T. Grance, "The NIST definition of cloud computing," Gaithersburg, MD, 2011. doi: 10.6028/NIST.SP.800-145.
- [2] F. Liu *et al.*, "NIST cloud computing reference architecture," Gaithersburg, MD, 2011. doi: 10.6028/NIST.SP.500-292.
- [3] A. Lackinger, A. Morichetta, and S. Dustdar, "Time Series Predictions for Cloud Workloads: A Comprehensive Evaluation," in *Proceedings - 18th IEEE International Conference on Service-Oriented System Engineering, SOSE 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1–10. doi: 10.1109/SOSE62363.2024.00007.
- [4] M. H. Berlian, R. Ferdiana, and G. D. Putra, "Proactive Cloud Sizing Automation for Platform-as-a-Service using Historical Workload Analysis," Institute of Electrical and Electronics Engineers (IEEE), Dec. 2025, pp. 374–381. doi: 10.1109/itis67966.2025.11308870.
- [5] K. Halder, U. Bellur, and P. Kulkarni, "Risk aware provisioning and resource aggregation based consolidation of virtual machines," in *Proceedings - 2012 IEEE 5th International Conference on Cloud Computing, CLOUD 2012*, 2012, pp. 598–605. doi: 10.1109/CLOUD.2012.86.
- [6] D. Saxena and A. K. Singh, "Workload Forecasting and Resource Management Models Based on Machine Learning for Cloud Computing Environments," Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2106.15112>
- [7] S. Bhagavathiperumal and M. Goyal, "Workload Analysis of Cloud Resources using Time Series and Machine Learning Prediction," 2019, doi: 10.1109/CSDE48274.2019.9162385.
- [8] T. Chen, R. Bahsoon, and X. Yao, "A Survey and Taxonomy of Self Aware and Self-Adaptive Cloud Autoscaling Systems," May 31, 2019, *Association for Computing Machinery*. doi: 10.1145/3190507.
- [9] I. K. Matthew, "Optimizing Cloud Computing Resources for Operational Cost and Application Performance Using Machine Learning," 2025.
- [10] M. H. Berlian, "A Systematic Review on Cloud Sizing Automation for PaaS Using Historical Workload Analysis," Nov. 20, 2025. doi: 10.36227/techrxiv.176366473.33006194/v1.
- [11] R. Aghili, Q. Qin, H. Li, and F. Khomh, "Understanding Web Application Workloads and Their Applications: Systematic Literature Review and Characterization," in *Proceedings - 2024 IEEE International Conference on Software Maintenance and Evolution, ICSME 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 474–486. doi: 10.1109/ICSME58944.2024.00050.
- [12] R. N. Calheiros, E. Masoumi, R. Ranjan, and R. Buyya, "Workload prediction using ARIMA model and its impact on cloud applications' QoS," *IEEE Transactions on Cloud Computing*, vol. 3, no. 4, pp. 449–458, 2015, doi: 10.1109/TCC.2014.2350475.
- [13] K. L. Devi and S. Valli, "Time series-based workload prediction using the statistical hybrid model for the cloud environment," *Computing*, vol. 105, no. 2, pp. 353–374, Feb. 2023, doi: 10.1007/s00607-022-01129-7.
- [14] J. Kumar, R. Goomer, and A. K. Signh, "Long Short-Term Memory Recurrent Neural Network Based Workload Forecasting Model," *International Conference on Smart Computing and Communications, ICSCC*, 2017.

- [15] Y. Zhu, W. Zhang, Y. Chen, and H. Gao, "A novel approach to workload prediction using attention-based LSTM encoder-decoder network in cloud environment," *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, no. 1, Dec. 2019, doi: 10.1186/s13638-019-1605-z.
- [16] R. Balamanigandan, R. Mahaveerakannan, S. Saraswathi, and A. M. Jenifer, "Optimizing Resource Costs With Machine Learning Techniques," in *2nd International Conference on Machine Learning and Autonomous Systems, ICMLAS 2025 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 364–370. doi: 10.1109/ICMLAS64557.2025.10968124.
- [17] V. Koc and V. Kamanuru, "Predictive and Reinforcement Learning-Based Framework for Cloud Resource Optimization," in *2025 IEEE Systems and Information Engineering Design Symposium, SIEDS 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 128–133. doi: 10.1109/SIEDS65500.2025.11021093.
- [18] H. Deshmukh, "Service Level Agreement -Driven Dynamic Resource Provisioning Strategy for Cloud Computing," in *3rd International Conference on Integrated Circuits and Communication Systems, ICICACS 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ICICACS65178.2025.10967993.
- [19] S. Saha, Affan, H. Raj, and A. U. Upadhyay, "ML-Driven VM Instance Selection for Cloud Workloads," in *Proceedings of 2025 International Conference on Emerging Technologies in Computing and Communication, ETCC 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ETCC65847.2025.11108372.
- [20] A. Khoshkbarforousha, R. Ranjan CSIRO, P. P. Jayaraman CSIRO, A. PremJayaraman, csiroau John Hosking, and E. Abbasnejad, "Resource Usage Estimation of Data Stream Processing Workloads in Datacenter Clouds."
- [21] F. Li, G. Wu, J. Lu, M. Jin, H. An, and J. Lin, "SmartCMP: A Cloud Cost Optimization Governance Practice of Smart Cloud Management Platform," in *Proceedings - 2022 IEEE 7th International Conference on Smart Cloud, SmartCloud 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 171–176. doi: 10.1109/SmartCloud55982.2022.00034.
- [22] S. K. Doddavula, I. Agrawal, and V. Saxena, "Cloud Computing Solution Patterns: Infrastructural Solutions," 2013, pp. 197–219. doi: 10.1007/978-1-4471-5107-4_10.
- [23] R. Ferdiana and G. D. Putra, *A Review of Cloud Migration Strategies in the Developing Country*. 4th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia, 2018.
- [24] R. B. Suryawan, R. Ferdiana, and Widyawan, "The Comparison of Cloud Migration Effort on Platform as a Service," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1742-6596/1577/1/012056.
- [25] M. P. Yadav, N. Pal, and D. K. Yadav, "Workload prediction over cloud server using time series data," in *Proceedings of the Confluence 2021: 11th International Conference on Cloud Computing, Data Science and Engineering*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 267–272. doi: 10.1109/Confluence51648.2021.9377032.
- [26] N. Mahesh and M. Supriya, "Auto-Scaling Anomaly Detection in Cloud Computing Models using MAPE-K Loop," in *2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICCCNT61001.2024.10725814.

- [27] A. Rossi, A. Visentin, D. Carraro, S. Prestwich, and K. N. Brown, "Forecasting Workload in Cloud Computing: Towards Uncertainty-Aware Predictions and Transfer Learning," Nov. 2023, [Online]. Available: <http://arxiv.org/abs/2303.13525>
- [28] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2018.
- [29] E. G. Radhika and G. S. Sadasivam, "A review on prediction based autoscaling techniques for heterogeneous applications in cloud environment," in *Materials Today: Proceedings*, Elsevier Ltd, 2021, pp. 2793–2800. doi: 10.1016/j.matpr.2020.11.789.
- [30] M. Abdullah, W. Iqbal, J. L. Berral, J. Polo, and D. Carrera, "Burst-Aware Predictive Autoscaling for Containerized Microservices," *IEEE Trans. Serv. Comput.*, vol. 15, no. 3, pp. 1448–1460, 2022, doi: 10.1109/TSC.2020.2995937.
- [31] M. Samir, K. T. Wassif, and S. H. Makady, "Proactive Auto-Scaling Approach of Production Applications Using an Ensemble Model," *IEEE Access*, vol. 11, pp. 25008–25019, 2023, doi: 10.1109/ACCESS.2023.3256302.
- [32] A. F. Baarzi and G. Kesidis, "SHOWAR: Right-sizing and efficient scheduling of microservices," in *SoCC 2021 - Proceedings of the 2021 ACM Symposium on Cloud Computing*, Association for Computing Machinery, Inc, Nov. 2021, pp. 427–441. doi: 10.1145/3472883.3486999.
- [33] M. K. Patra, B. Sahoo, A. K. Turuk, and S. Misra, "Task grouping and optimized deep learning based VM sizing for hosting containers as a service," *Journal of Cloud Computing*, vol. 12, no. 1, Dec. 2023, doi: 10.1186/s13677-023-00441-7.
- [34] L. Krautsevich, A. Lazouski, F. Martinelli, and A. Yautsiukhin, "Risk-aware usage decision making in highly dynamic systems," in *5th International Conference on Internet Monitoring and Protection, ICIMP 2010*, 2010, pp. 29–34. doi: 10.1109/ICIMP.2010.13.
- [35] L. Jiang and Y. Wang, "Risk-aware Decision-making in Human-multi-robot Collaborative Search: A Regret Theory Approach," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 105, no. 2, Jun. 2022, doi: 10.1007/s10846-022-01642-z.
- [36] T. Llorido-Botran, J. Miguel-Alonso, and J. A. Lozano, "A Review of Auto-scaling Techniques for Elastic Applications in Cloud Environments," *J. Grid Comput.*, vol. 12, no. 4, pp. 559–592, Nov. 2014, doi: 10.1007/s10723-014-9314-7.
- [37] M. Schultzberg, S. Ankargren, and M. Frånberg, "Risk-aware product decisions in A/B tests with multiple metrics," *J. Stat. Plan. Inference*, vol. 245, p. 106393, Feb. 2026, doi: 10.1016/j.jspi.2026.106393.
- [38] N. Roy, A. Dubey, and A. Gokhale, "Efficient autoscaling in the cloud using predictive models for workload forecasting," in *Proceedings - 2011 IEEE 4th International Conference on Cloud Computing, CLOUD 2011*, 2011, pp. 500–507. doi: 10.1109/CLOUD.2011.42.
- [39] M. R. R. Deva and N. Jain, "Utilizing Azure Automated Machine Learning and XGBoost for Predicting Cloud Resource Utilization in Enterprise Environments," in *2025 International Conference on Networks and Cryptology, NETCRYPT 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 535–540. doi: 10.1109/NETCRYPT65877.2025.11102235.
- [40] R. Scholar and S. Jayalakshmi, "Predictive Scaling for Elastic Compute Resources on Public Cloud Utilizing Deep Learning based Long Short-term Memory." [Online]. Available: www.ijacsa.thesai.org

- [41] P. Osypanka and P. Nawrocki, "Resource Usage Cost Optimization in Cloud Computing Using Machine Learning," *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 2079–2089, 2022, doi: 10.1109/TCC.2020.3015769.
- [42] P. Malathi, E. Dilipkumar, J. Rajasubha, A. S. Yokesh, L. Priya, and H. A. Basha, "Optimizing Resource Allocation in Cloud Environments Using Reinforcement Learning," in *Proceedings of the 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICSES63760.2024.10910650.
- [43] L. M. Al Qassem, T. Stouraitis, E. Damiani, and I. A. M. Elfadel, "Proactive Random-Forest Autoscaler for Microservice Resource Allocation," *IEEE Access*, vol. 11, pp. 2570–2585, 2023, doi: 10.1109/ACCESS.2023.3234021.
- [44] D. Buchaca, J. Ll. Berral, C. Wang, and A. Youssef, "Proactive container auto-scaling for cloud native machine learning services," in *IEEE International Conference on Cloud Computing, CLOUD*, IEEE Computer Society, Oct. 2020, pp. 475–479. doi: 10.1109/CLOUD49709.2020.00070.
- [45] V. K. Jayakumar, S. Arbat, I. K. Kim, and W. Wang, "CloudBruno: A Low-Overhead Online Workload Prediction Framework for Cloud Computing," in *Proceedings - 2022 IEEE International Conference on Cloud Engineering, IC2E 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 188–198. doi: 10.1109/IC2E55432.2022.00027.
- [46] W. Iqbal, A. Erradi, M. Abdullah, and A. Mahmood, "Predictive Auto-Scaling of Multi-Tier Applications Using Performance Varying Cloud Resources," *IEEE Transactions on Cloud Computing*, vol. 10, no. 1, pp. 595–607, 2022, doi: 10.1109/TCC.2019.2944364.
- [47] N. Kumawat, N. Handa, and A. Kharbanda, "Cloud Computing Resources Utilization and Cost Optimization for Processing Cloud Assets," in *Proceedings - 2020 IEEE International Conference on Smart Cloud, SmartCloud 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 41–48. doi: 10.1109/SmartCloud49737.2020.00017.
- [48] G. Baghel, P. Singh, and P. Garg, "Dynamic Resource Allocation Strategies for Cloud Computing: A User-Centric Approach Using AI," in *ICDT 2025 - 3rd International Conference on Disruptive Technologies*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 1267–1270. doi: 10.1109/ICDT63985.2025.10986665.
- [49] J. Tyagi and M. Sharma, "Optimising Resource Sharing Algorithms for Efficient Cloud Computing," in *Proceedings - 2025 IEEE 11th Conference on Big Data Security on Cloud, BigDataSecurity 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 46–51. doi: 10.1109/BigDataSecurity66063.2025.00028.
- [50] R. David, D. K. Arivalagan, A. S. Prabu, S. Ravi, P. Giri, and A. A. A. Samhan, "Optimizing Cloud Computing Efficiency for Scalable and Adaptive Resource Management via AI and Machine Learning," in *2025 Global Conference in Emerging Technology, GINOTECH 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/GINOTECH63460.2025.11076855.
- [51] S. T. Singh, M. Tiwari, and A. S. Dhar, "Machine Learning based Workload Prediction for Auto-scaling Cloud Applications," in *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development, OTCON 2022*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/OTCON56053.2023.10114033.

- [52] S. Ambi Karthikeyan, "Azure Automation Runbook Types," in *Azure Automation Using the ARM Model*, Apress, 2017, pp. 59–86. doi: 10.1007/978-1-4842-3219-4_3.
- [53] P. Osypanka and P. Nawrocki, "QoS-Aware Cloud Resource Prediction for Computing Services," *IEEE Trans. Serv. Comput.*, vol. 16, no. 2, pp. 1346–1357, Mar. 2023, doi: 10.1109/TSC.2022.3164256.
- [54] G. Mudali, D. S. Roy, and K. H. K. Reddy, "QoS Aware Heuristic Provisioning Approach for Cloud Spot Instances," in *Proceedings - 2017 International Conference on Information Technology, ICIT 2017*, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 73–78. doi: 10.1109/ICIT.2017.59.
- [55] S. Mireslami, L. Rakai, M. Wang, and B. H. Far, "Dynamic Cloud Resource Allocation Considering Demand Uncertainty," *IEEE Transactions on Cloud Computing*, vol. 9, no. 3, pp. 981–994, Jul. 2021, doi: 10.1109/TCC.2019.2897304.
- [56] Mohammad Robihul Mufid, Arif Basofi, and M. Udin Harun Al Rasyid, *Design an MVC Model using Python for Flask Framework Development*. International Electronics Symposium (IES), 2019.
- [57] Sasalak Tongkaw and Aumnat Tongkaw, *A Comparison of Database Performance of MariaDB and MySQL with OLTP Workload*. IEEE Conference on Open Systems (ICOS), October 10-12, 2016, Langkawi, Malaysia, 2016.
- [58] P. Bhardwaj, "Driving Operational Efficiency through Azure Automation and PowerShell," 2021.
- [59] Microsoft, "Autoscale in Azure Monitor." Accessed: May 16, 2026. [Online]. Available: <https://learn.microsoft.com/en-us/azure/azure-monitor/autoscale/autoscale-overview>
- [60] F. Al-Haidari, M. Sqalli, and K. Salah, "Impact of CPU utilization thresholds and scaling size on autoscaling cloud resources," in *Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom*, IEEE Computer Society, 2013, pp. 256–261. doi: 10.1109/CloudCom.2013.142.
- [61] M. K. M. Murthy, H. A. Sanjay, and J. Anand, "Threshold Based Auto Scaling of Virtual Machines in Cloud Environment," 2014.
- [62] E. Casalicchio, "A study on performance measures for auto-scaling CPU-intensive containerized applications," *Cluster Comput.*, vol. 22, no. 3, pp. 995–1006, Sep. 2019, doi: 10.1007/s10586-018-02890-1.
- [63] Microsoft, "DTU-based purchasing model - Azure SQL Database." Accessed: May 16, 2026. [Online]. Available: <https://learn.microsoft.com/en-us/azure/azure-sql/database/service-tiers-dtu?view=azuresql>
- [64] Microsoft, "Monitoring Azure SQL Database with metrics and alerts." Accessed: May 16, 2026. [Online]. Available: <https://learn.microsoft.com/en-us/azure/azure-sql/database/monitoring-metrics-alerts>