

Predicting Cloud Resource Provisioning using Machine Learning Techniques

Prasad Madhukarrao Nakhate¹, Vrushabh Vinod Sasane², Vedant Rajendra Sanap³,
Siddhi Ramchandra Kore⁴

UG Students, Department of Computer Engineering^{1,2,3,4}
Sinhgad College of Engineering, Pune, Maharashtra, India

Abstract: *It can be difficult to provision assets in a cloud environment. Both inadequate and excessive provisioning are bad for the system's overall performance. Researchers are increasingly turning to proactive provisioning techniques that foresee resource needs in advance and set the system up to accommodate such real-time demands. Although proactive provisioning methods are more complicated than reactive provisioning strategies, they offer a generally faster reaction time since decisions about resource provisioning are made in advance of the actual requirement for those resources. The adoption of an analytical framework that anticipates resource requirements is necessary for these proactive provisioning strategies to function well. The provisioning of resources for computational operations in cloud computing is a significant concern. Unsurprisingly, previous research has identified a number of ways to deliver Cloud resources effectively. However, a forecast of the impending computational jobs' future resource consumption is necessary to implement a comprehensive resource provisioning model is required. However, the subject of cloud prediction the use of resources is still in its early stages.*

Keywords: Proactive provisioning, Resource prediction, Prediction model, Machine learning, KNN, SVM, RF.

REFERENCES

- [1]. Iglesias, J.O., et al. A Methodology for Online Consolidation of Tasks through More Accurate Resource Estimations. in Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on. 2014.
- [2]. Caglar, F. and A. Gokhale. iOverbook: Intelligent Resource-Overbooking to Support Soft Real-Time Applications in the Cloud. in Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on. 2014.
- [3]. Qi, Z., et al. Harmony: Dynamic Heterogeneity-Aware Resource Provisioning in the Cloud. in Distributed Computing Systems (ICDCS), 2013 IEEE 33rd International Conference on. 2013.
- [4]. Hu, R., et al., Efficient Resources Provisioning Based on Load Forecasting in Cloud. The Scientific World Journal, 2014. 2014: p. 12.
- [5]. Reiss, C., A. Tumanov, and G. Ganger, Towards understanding heterogeneous clouds at scale: Google trace analysis. Center for cloud computing, 2012.
- [6]. Kundu, S. et al., "Modeling virtualized applications using machine learning techniques", in Proc. Of 8th ACM SIGPLAN /Sigpos conference on Virtual Execution Environments, pp3 – 14, London, UK 2012
- [7]. Kupferman, J. et al., "Scaling Into the Cloud". University Of California, Santa Barbara, Tech. Rep. <http://cs.ucsb.edu/~jkupferman/docs/ScalingIntoTheCloud s.pdf>. 2009.
- [8]. Quiroz, A et al., "Towards autonomic workload provisioning for enterprise Grids and clouds" in Grid Computing, 2009 10th IEEE/ACM International Conference pp 50-57, October, 2009
- [9]. Sadeka, I. et al., "Empirical prediction models for adaptive resource provisioning in the cloud", Future Generation Computer Systems, vol. 28, no. 1, pp 155 – 165, January, 2012
- [10]. Sakr, G.E et al., "Artificial intelligence for forest fire prediction" IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp.1311-1316, July 2010.

- [11]. Sapankevych, N and Sankar, R., "Time Series Prediction Using Support Vector Machines: A Survey," Computational Intelligence Magazine, IEEE, vol.4, no.2, pp.24-38, May 2009.
- [12]. H. Wang, H. Shen, and Z. Li, "Approaches for resilience against cascading failures in cloud datacenters," in *Proc. of ICDCS*, 2018.
- [13]. W. Wei, H. Fan, X. and Song, and J. Fan, X. and Yang, "Imperfect information dynamic stackelberg game based resource allocation using hidden markov for cloud computing," *Trans. on SC*, 2018.
- [14]. M. Xu and R. Buyya, "Brownout approach for adaptive management of resources and applications in cloud computing systems: A taxonomy and future directions," *ACM Computing Surveys (CSUR)*, 2019.
- [15]. Y. Yu, F. Jindal, V. and Bastani, F. Li, and I. Yen, "Improving the smartness of cloud management via machine learning based workload prediction," in *Proc. of COMPSAC*, 2018.
- [16]. M. Hassan, H. Chen, and Y. Liu, "Dears: A deep learning based elastic and automatic resource scheduling framework for cloud applications," in *Proc. of UBIComp*, 2018.