IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

Optimizing Service Placement and Enhancing Service Allocation for Microservice Architectures in Cloud Environments

Roshan Mahant[1], Sumit Bhatnagar[2], Vikas Mendhe[3]
Launch IT Corp, Urbandale, IA USA[1,3]
JP Morgan Chase & Co., New Jersey, USA[2]

Abstract: *With the increasing popularity of microservice architecture, there is a growing need to deploy service-based applications efficiently in cloud environments. Traditional cluster schedulers often fail to optimize service placement adequately, as they only consider resource constraints and overlook traffic demands between services. This oversight can lead to performance issues such as high response times and jitter. To address this challenge, we propose a novel approach to optimize the placement of service-based applications in clouds. Our approach involves partitioning the application into segments while minimizing overall traffic between them, and then strategically allocating these segments to machines based on their resource and traffic demands. We have developed a prototype scheduler and conducted extensive experiments on test bed clusters to evaluate its performance. The results demonstrate that our approach surpasses existing container cluster schedulers and heuristic methods, significantly reducing overall inter-machine traffic and improving application performance.*

Keywords: Microservice architecture, Cloud computing, Service-based applications

DOI: 10.48175/IJARSCT-18276