

A Cost-Efficient Container Orchestration Strategy in Kubernetes-Based Cloud Computing Infrastructures with Heterogeneous Resources

ZHIHENG ZHONG and RAJKUMAR BUYYA, University of Melbourne

Containers, as a lightweight application virtualization technology, have recently gained immense popularity in mainstream cluster management systems like Google Borg and Kubernetes. Prevalently adopted by these systems for task deployments of diverse workloads such as big data, web services, and IoT, they support agile application deployment, environmental consistency, OS distribution portability, application-centric management, and resource isolation. Although most of these systems are mature with advanced features, their optimization strategies are still tailored to the assumption of a static cluster. Elastic compute resources would enable heterogeneous resource management strategies in response to the dynamic business volume for various types of workloads. Hence, we propose a heterogeneous task allocation strategy for cost-efficient container orchestration through resource utilization optimization and elastic instance pricing with three main features. The first one is to support heterogeneous job configurations to optimize the initial placement of containers into existing resources by task packing. The second one is cluster size adjustment to meet the changing workload through autoscaling algorithms. The third one is a rescheduling mechanism to shut down underutilized VM instances for cost saving and reallocate the relevant jobs without losing task progress. We evaluate our approach in terms of cost and performance on the Australian National Cloud Infrastructure (Nectar). Our experiments demonstrate that the proposed strategy could reduce the overall cost by 23% to 32% for different types of cloud workload patterns when compared to the default Kubernetes framework.

CCS Concepts: • **Computer systems organization** → **Cloud computing**; • **Theory of computation** → **Scheduling algorithms**; • **General and reference** → Performance; • **Computing methodologies** → Model development and analysis;

Additional Key Words and Phrases: Cluster management, container orchestration, resource heterogeneity, cost efficiency

ACM Reference format:

Zhiheng Zhong and Rajkumar Buyya. 2020. A Cost-Efficient Container Orchestration Strategy in Kubernetes-Based Cloud Computing Infrastructures with Heterogeneous Resources. *ACM Trans. Internet Technol.* 20, 2, Article 15 (April 2020), 24 pages.
<https://doi.org/10.1145/3378447>

This work was supported by the China Scholarship Council and the Australia Research Council Discovery Project.

Authors' address: Z. Zhong and R. Buyya, University of Melbourne, Cloud Computing and Distributed Systems (CLOUDS) Laboratory, School of Computing and Information System, Parkville Campus, Melbourne, Victoria, 3010, Australia; emails: zhiheng@student.unimelb.edu.au, rbuyya@unimelb.edu.au.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

1533-5399/2020/04-ART15 \$15.00

<https://doi.org/10.1145/3378447>