

마이크로서비스 아키텍처의 배포 비용을 최적화하는 알고리즘

리즈양⁰, Scott Uk-Jin Lee^{*}

⁰한양대학교 컴퓨터공학과,

^{*}한양대학교 컴퓨터공학과

e-mail: {liziangxy⁰, scottle^{*}}@hanyang.ac.kr

An Algorithm to Optimize Deployment Cost for Microservice Architecture

Ziang Li⁰, Scott Uk-Jin Lee^{*}

⁰Dept. of Computer Science and Engineering, Hanyang University,

^{*}Dept. of Computer Science and Engineering, Hanyang University

● 요약 ●

As the utilization of microservice architectural style in diverse applications are increasing, the microservice deployment cost became a concern for many companies. We propose an approach to reduce the deployment cost by generating an algorithm which minimizes the cost of basic operation of a physical machine and the cost of resources assigned to a physical machine. This algorithm will produce optimal resource allocation and deployment location based on genetic algorithm process.

키워드: 마이크로서비스(microservice), 배포 비용(deployment cost), 자원 할당(resource allocation)

I. Introduction

Microservice providers bear all costs of data center, making cost of deploying microservice systems an important factor to be considered. In this research, we propose an algorithm based approach to minimize the cost of microservice deployment. We have obtained the global optimal solution with a genetic algorithm method that can produce optimal resource allocation and deployment location.

and finally adopt genetic algorithm. There are various types of user requests, the execution process of each request is represented as a directed acyclic graph shown in Fig. 1 below.

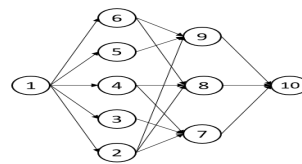


Fig. 1. Microservice Business Logic

II. Related Work

The FreeContainer method[1] reduces costs by reassigning containers. Converting container deployment to an Integer Linear Programming issue[2] minimizes deployment costs. However, these related works treat microservice as an individual without considering the sharing characteristic of images on containers.

1. Objective function

The total quantity of physical machines is N , the cost of basic running of n^{th} machine is $P(n)$, and the usage cost of computational resources per unit is P_0 . We assume that K kinds of images are required for services. The cost of basic operation of the n physical machine is as follows:

$$Acost = P(n) \times s(n)$$

III. The Proposed Scheme

We first design the execution process of microservice requests, determine variables, define the objective function and constraints,

$$s(n) = \begin{cases} 1, & \sum_{k=1}^K l(k,n) \neq 0 \\ 0 & \text{other} \end{cases}$$

The quantity of microservices is M where the m^{th} microservice has $D(m)$ instructions, the maximum time of each microservice completing a request is denoted as $T(m)$. R types of requests are satisfied in microservice business logic. The number of user requests from the r^{th} kind of request to the m^{th} microservice is denoted as $E(r,m)$. The minimum total computational resource required for each microservice can be calculated as below.

$$C(m) = \frac{\left[\sum_{r=1}^R E(r,m) \right] \times D(m)}{T(m)}$$

We assume multiple images are required for each service and denoted the number of images required for the m^{th} service as $I(m)$. The cost of resources assigned to the n physical machine is as follows:

$$Bcost = \sum_{m=1}^M \frac{\sum_{r=1}^R x(r,m,I(m),n)}{E(m)} \times P_0(n) \times C(m)$$

Based on the above definitions, the objective function can be defined as minimizing the sum of the above two costs.

$$\min \left\{ \sum_{n=1}^N [Acost + Bcost] \right\}$$

2. Constraints

To make sure computational resources of each physical machines are not wasted, some constraints need to be satisfied when deploying microservice.

Constraint 1: Each physical machine handles a large number of requests. All requests are executed.

$$\sum_{r=1}^R \sum_{m=1}^M x(r,m,I(m),n) = E(m), \quad \forall m \in M$$

Constraint 2: To ensure rationality of allocation, each physical machine should not have service duplications. Only one service of the same type can be deployed in a physical machine.

$$\sum_{r=1}^R \sum_{m=1}^M \frac{x(r,m,I(m),n)}{x(r,m,I(m),n)} \leq M, \quad \forall n \in N$$

Constraint 3: There are also some ranges of values that we need to set.

$$l(k,n) \in \{0,1\}, \quad x(r,m,I(m),n) \in \{0,1,2,\dots,E(m)\}$$

3. Optimization

We have applied genetic algorithm method to the above objective function to gain the global optimal solution under

constraints defined above. The results show that we can obtain optimal resource allocation and deployment location of microservice. For different initial population, the global optimal solution can always be found.

IV. Conclusions

In this paper, we proposed an algorithm, which minimizes the cost of the running and allocated resources of a physical machine to optimize the cost for microservice deployment while considering the cost of microservice images. The optimal resource allocation and deployment location can be obtained by genetic algorithm method to provide higher efficiency and lower costs when deploying microservice systems.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2019R1H1A1101270).

REFERENCES

- [1] Y. Zhang, Y. Li, K. Xu, D. Wang, M. Li, X. Cao, Q. Liang, "A Communication-Aware Container Re-Distribution Approach for High Performance VNFs," 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), IEEE, 2017.
- [2] D. Zhang, B. Yan, Z. Feng, C. Zhang, Y. Wang, "Container oriented job scheduling using linear programming model," 2017 3rd International Conference on Information Management (ICIM), IEEE, 2017.