Task Allocation Scheme for Hadoop in Campus Network Environment

Tomohiro Matsuno† Bijoy Chand Chatterjee† Satoru Okamoto† Eiji Oki† Malathi Veeraraghavan‡ Naoaki Yamanaka*
†Dept. of Communication Engineering and Informatics, The University of Electro-Communications, Japan
‡Dept. of Electrical and Computer Engineering, University of Virginia, USA
*Dept. of Information and Computer Science, Keio University, Japan

I. INTRODUCTION

Different types of available computational resources can be utilized with the incorporation of Hadoop framework [1], [2]. In the current Hadoop framework [1], [3], the job is divided into the same proportion of tasks and they are assigned to slave servers. If the performance of computational resources is different, the processing time cannot be decreased, even if the number of computational resources is increased. In the other words, when low performance computational resources are included, the advantages of Hadoop cannot be utilized properly. As a result, we cannot utilize the outdated resources in campus environments with incorporation of the current Hadoop framework.

This paper proposes a resource allocation scheme for Hadoop in campus networks in order to maximize the performance in terms of processing time by utilizing the outdated computational resources. The proposed scheme splits the job with appropriate splitting ratios according to both computer processing performance and network resource availability. In the other words, the job is divided into different proportion of tasks and then they are assigned to slave servers according to both computer processing performance and network resource availability. The simulation results indicate that the proposed scheme is effective.

II. PROPOSED SCHEME

The proposed scheme is intended to maximize the performance in terms of processing time by utilizing the outdated computational resources in campus networks. If we use multiple client servers with different performance, splitting the job with appropriate splitting ratios according to both computer processing performance and network resource availability suppresses the overall processing time. The processing time is suppressed by minimizing the maximum processing time of slave servers. The proposed scheme determines the appropriate splitting ratios according to computer processing performance. The concept of the proposed scheme, which splits the job with appropriate splitting ratios according to both computer processing performance and network resource availability and assigns tasks of different proportion to slave servers is shown in Fig. 1.

III. PERFORMANCE EVALUATION

In our simulation, we consider 18 computers as slave servers; out of them, nine computers are modern with high performance, and remaining nine computers are outdated with low performance. Figure 2 shows the normalized processing time of \( r \), which is maximum processing time among all tasks, versus the number of used slave servers. The processing time of \( r \) is normalized by that of one slave server. We can observe that the normalized processing time of \( r \) using the proposed scheme decreases with increase in slave servers. This is due to the appropriate assignment of splitting ratio of the job to computational resources. On the other hand, it is observed that the normalized processing time of \( r \) using the conventional scheme is increased, when the number of slave servers is considered greater than nine.

In the other words, we can say that the conventional scheme does not perform well when the poor performance computers are included as slave servers. From the above discussion, we conclude that the proposed scheme utilizes the outdated poor performance computational resources and suppresses the processing time compared to the conventional scheme.

IV. CONCLUSION

This paper proposed a Hadoop resource allocation scheme in campus networks that maximizes performance in terms of processing time by utilizing the outdated computational resources. The proposed scheme splits a computational job into tasks using appropriate splitting ratios based on computer processing performance and network resource availability. We compared our resource allocation scheme with a conventional scheme, which divides a job equally without consideration of computer processing performance and network resource availability. The simulation results indicated that the proposed scheme outperforms the conventional scheme in terms of processing time.

ACKNOWLEDGEMENT

This work was supported in part by the National Institute of Information and Communications Technology (NICT), Japan and the NSF grant (CNS-1405171), USA.

REFERENCES