Abstract

Parallel data processing has become more and more reliable phenomenon due to the realization of cloud computing, especially using IaaS (Infrastructure as a Service) clouds. The cloud service providers such as IBM, Google, Microsoft and Oracle have made provisions for parallel data processing in their cloud services. Nevertheless, the frameworks used as of now are static and homogenous in nature in a cluster environment. The problem with these frameworks is that the resource allocation when large jobs are submitted is not efficient as they take more time for processing besides incurring more cost. In this paper we discuss the possibilities of parallel processing and its challenges. One of the IaaS products meant for parallel processing is presented in this paper. VMs are allocated to tasks dynamically for execution of jobs. With proposed framework we performed parallel job processing which involves Map Reduce, a new programming phenomenon. We also compare this with Hadoop.

References

- J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large


- G. von Laszewski, M. Hategan, and D. Kodeboyina. Workflows for e-Science Scientific
Parallel Data Processing for Effective Dynamic Resource Allocation in the Cloud

- D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, and D. Zagorodnov. Eucalyptus: A

Index Terms
Computer Science
Information Sciences

Keywords
Parallel processing cloud computing Map Reduce many-task computing