Indonesian Journal of Electrical Engineering and Computer Science
Vol. 30, No. 2, May 2023, pp. 1059~1067
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v30.i2.pp1059-1067

An innovativefractal architecture model for implementing MapReduce in an open multiprocessing parallel environment

Muslim Mohsin Khudhair[1,3], Adil Al-Rammahi[2], Furkan Rabee[3]
[1]Department of Computer Information System, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq
[2]Department of Mathematical Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq
[3]Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

Article Info

Article history:

Received Nov 11, 2022
Revised Dec 25, 2022
Accepted Dec 28, 2022

Keywords:

Cloud computing
Cube model
Fractal
MapReduce
Open multiprocessing
Shared-memory

ABSTRACT

One of the infrastructure applications that cloud computing offers as a service is parallel data processing. MapReduce is a type of parallel processing used more and more by data-intensive applications in cloud computing environments. MapReduce is based on a strategy called "divide and conquer," which uses regular computers, also called "nodes," to do processing in parallel. This paper looks at how open multiprocessing (OpenMP), the best shared-memory parallel programming model for high-performance computing, can be used with the proposed fractal network model in the MapReduce application. A well-known model, the cube, is used to compare the fractal network model and its work. Where experiments demonstrated that the fractal model is preferable to the cube model. The fractal model achieved an average speedup of 2.7 and an efficiency rate of 67.7%. In contrast, the cube model could only reach an average speedup of 2.5 and an efficiency rate of 60.4%.

Corresponding Author:

Muslim Mohsin Khudhair
Department of Computer Information System, College of Computer Science and Information Technology
University of Basrah
Basrah, Iraq
Email: muslim.khudhair@uobasrah.edu.iq

1. INTRODUCTION

Cloud computing is a new technology growing quickly in the information technology (IT) industry. It uses different ideas and technologies, such as virtualization, processing power, storage, sharing, distributed networks, and connectivity, to its advantage. Cloud-based services are now one of the best ways for users and businesses to get on-demand services and unlimited storage [1]. One of the most important things about a cloud-based platform is that it makes it easy to quickly process large data sets in cloud applications. To deploy big data applications on a cloud data centre network (DCN), the key challenges are volume, velocity, and variety. Volume is the amount of data, variety is the number of data types, and velocity is the speed at which data is processed [2], [3].

Big data is becoming an increasingly important way for a business to enhance its value proposition or the efficiency of its operations. Due to the sheer amount of data, big data relies heavily on parallel computing technology to finish processing data quickly. Many specialized programming models and runtime systems have been created to support big data [4]. The map/reduce approach is utilized by Hadoop and Spark [5]. GraphLab, Giraph, and GraphX use the Pregel model [6]. Storm can deal with flowing data [7]. Each system uses threads