Fragment Re-Allocation Strategy Based on Hypergraph for NoSQL Database Systems

Zhikun Chen, Unit 91655, People’s Liberation Army, Beijing, China
Shuqiang Yang, Department of Computer Science, National University of Defense Technology, Changsha, China
Yunfei Shang, Unit 91655, People’s Liberation Army, Beijing, China
Yong Liu, Unit 91655, People’s Liberation Army, Beijing, China
Feng Wang, Unit 91655, People’s Liberation Army, Beijing, China
Lu Wang, Unit 91655, People’s Liberation Army, Beijing, China
Jingjing Fu, Unit 95025, People’s Liberation Army, Beijing, China

ABSTRACT

NoSQL database is famed for the characteristics of high scalability, high availability, and high fault-tolerance. It is used to manage data for a lot of applications. The computing model has been transferred to “computing close to data”. Therefore, the location of fragment directly affects system’s performance. Every site’s load dynamical changes because of the increasing data and the ever-changing operation pattern. So system has to re-allocate fragment to improve system’s performance. The general fragment re-allocation strategies of NoSQL database scatter the related fragments as possible to improve the operations’ parallel degree. But those fragments may interact with each other in some application’s operations. So the high parallel degree of operation may increase system’s communication cost such as data are transferred by network. In this paper, the authors propose a fragment re-allocation strategy based on hypergraph. This strategy uses a weighted hypergraph to represent the fragments’ access pattern of operations. A hypergraph partitioning algorithm is used to cluster fragments in the strategy. This strategy can improve system’s performance according to reducing the communication cost while guaranteeing the parallel degree of operations. Experimental results confirm that the strategy will effectively contribute in solving fragment re-allocation problem in specific application environment of NoSQL database system, and it can improve system’s performance.

KEYWORDS

Clustering Fragments, Fragment Allocation, Fragment Correlation, Hypergraph Partition, NoSQL Database

1. INTRODUCTION

With the development of Internet and Cloud Computing (S.K., 2007) technologies, the world has been in the information explosion era. For example, the number of internet’s users and connected devices both rapidly grew in the decade years. So every application system has to face the challenge of big data (Big data, 2013) which is a hot field of academia and industrial community recently. The big data has been discussed in literatures (Nature, 2008; Bryant et al., 2008; science.com, 2011; Manyika et al., 2011; W. E. Forum, 2012; Economist, 2010, 2011; Lohr, 2012; Noguchi, 2011; Whitehouse,
2012) which involve a lot of fields such as government strategy, medicine, and economic strategy and so on. Under the big data environment, the data of the application system have the following characteristics (B.H., 2011): Volume (the scale of data and users are both very large), Variety (the data source and data type are both variety), Velocity (the increment speed of data is fast), and Value (the data has a lot of potential value). Under the big data era, how to store and manage the big data is a big problem for those application systems. The traditional database technologies can not satisfy the requirement of high concurrent users and operations. The emergence of NoSQL Database (B.H., 2013) provides a good solution for those applications’ data storage and management.

NoSQL database is a new data management system, which is generated under the background of cloud computing. The computing model of big data environment transfers from “data close to computing” to “computing close to data as possible as it can”. The location of fragment will directly affect system’s performance. Under the big data era, the load of sites dynamical changes because of the system’s ever-changing operation pattern. So we have to dynamically regulate fragment’s location to guarantee system’s performance. The general fragment re-allocation strategies of NoSQL database scatter the related fragments as possible, which can improve the parallel degree of operation. But those fragments may interact with each other in some applications’ operations. The high parallel degree of operations may increase the communication cost such as data are transferred by network. Kumar et al. (2013) had used some experiments to prove this issue in the distributed database. In this paper, the objective of our fragment re-allocation strategy is to reduce the total resource consumption according to reducing the number of machine involved in the execution of an operation (called operation span). In order to reduce the operation span, we should allocate the fragments with high relative to the same site as possible while guaranteeing the parallel degree of operation. We use a hypergraph model to represent the relative of fragment. A hypergraph partition algorithm will be used to cluster fragments, which has to cluster the relative fragments together while guaranteeing parallel degree of operation. Finally, a fragments cluster allocation algorithm will be proposed.

The contributions of this paper are as follows:

1. We analyze and research the fragment allocation strategy of NoSQL database in the big data environment in detail;
2. We propose a new effective Fragment re-Allocation Strategy Based on Hypergraph (FASBH) and allocation algorithm for some specific application. The strategy can reduce the system’s communication cost while guaranteeing the parallel degree of operations;
3. Finally, we use several experiments to verify the effectiveness and performance of this strategy.

The rest of this paper is organized as follows. Section 2 introduces some preliminary knowledge of this paper. First, it describes the problem being tackled in this paper and some definitions. Section 3 states the hypergraph model and the implement of FASBH in detail. In section 4, we use some experiments to verify the effectiveness and performance of FASBH. We also compare the performance of strategies based on hypergraph partition model with strategies based on graph partition model. Section 5 describes the related work of fragment allocation in the database system (NoSQL and Distributed database). Section 6 concludes this paper.

2. PRELIMINARIES

This section introduces some preliminary knowledge, such as definition, problem state, and hypotheses and so on. Some definitions are introduced first, after that the problem state will be described. Some hypothesis will be shown in the next. The difference of fragment allocation problem between the NoSQL database and Distributed database will be stated at last.
User Interface Distribution Method Based on Pi-Calculus
www.igi-global.com/article/user-interface-distribution-method-based-on-pi-calculus/232303?camid=4v1a

Scalable Index and Data Management for Unstructured Peer-to-Peer Networks
www.igi-global.com/chapter/scalable-index-data-management-unstructured/36406?camid=4v1a