Skipnet-Octree Based Indexing Technique for Cloud Database Management System

Shweta Malhotra, Jamia Millia Islamia University, New Delhi, India
Mohammad Najmud Doja, Department of Computer Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia University, New Delhi, India
Bashir Alam, Department of Computer Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia University, New Delhi, India
Mansaf Alam, Big Data and Cloud Computing Laboratory, Department of Computer Science, Jamia Millia Islamia University, New Delhi, India

ABSTRACT

This article describes how data indexing plays a very crucial role in query processing. Systems based on traditional indexes like B-tree, R-tree, Bitmap, inverted indexing techniques are not suitable for efficient query evaluation as these systems are based on simple key-value pair and used only for point queries. In cloud data repositories, point queries are not sufficient for query as a cloud consists of multidimensional data. For multidimensional query processing, many techniques have been developed. In this article, a dynamic double layer indexing structure with the help of a Skipnet overlay for global indexing and an Octree index technique for local indexing has been proposed. It has been concluded from the experiments that Skipnet-Octree performs better than the previous double-layer indexing technique for complex queries.

KEYWORDS
Cloud Database Indexing Technique, Multidimensional Index, Octree, Skipnet, SkipNet-Octree

1. INTRODUCTION

Recently, cloud computing services are in demand because they provide inexpensive and powerful resources to the users (Armirust et al., 2010). For managing cloud database, cloud storage services plays a very important role. For efficient accessing of these data sets indexing plays a very crucial role. Current systems based on Distributed hash table works well with key-value pairs to give best performance for exact matching through point queries. It does not support or provide good performance for multidimensional and complex queries.

Many advanced indexing strategies such as R tree, Quad Tree, Bitmap, Octree has been developed with many advantages and limitations for efficiently accessing multidimensional queries. Octree is considered suitable for multi-attribute data and multidimensional queries. 2014, a two level SNB-Index based on SkipNet and B+Tree is proposed (Zhou et al., 2014) which provides high availability and scalability and support range queries as well as similarity queries but the proposed model was efficient for single attribute queries not for multi-attribute queries. In (2015, 2016) a decentralized, scalable double layer indexing structure using probabilistic Skiplist and Octree has been proposed (Dong et al., 2015) (He et al., 2016), skip-list is used to speed up the searching process and provides hierarchical querying and support multidimensional query using Octree indexing technique. A Skiplist (Pugh,
1990) is an in-memory data structure which can be used in place of balanced trees. Instead of forced balancing algorithm randomized algorithm is used for balancing which gives better results. With the help of probabilistic pattern, insertion and deletion operations are faster as compared to balanced trees. While processing any query not all elements of the data are useful hence Skiplist provides a way to skip many data element as it is based on hierarchical ordered linked list. In Skiplist traversing is done only from the head node whereas SkipNet (Harvey et al., 2003) provides a facility to traverse from any node. In Skiplist, each data node maintains variable number of pointer and different amount of traversal traffic per data record. But in case of SkipNet each node maintains 2LogN pointers where N refers to the number of nodes in the form of routing table. Since these nodes are used to pass message traffic in between nodes with the help of these routing table.

SkipNet (Harvey et al., 2003) is distributed generalization of Skiplist with two advantages i.e. meticulous data placement and data guaranteed routing locality. SkipNet works based on these two locality properties content locality and path locality. Content locality refers to dynamically or distributed placement of data on the overlay nodes within the range of organization and path locality means if a message is passed between two overlay networks it should be routed within the same organization. SkipNet is double ring linked list so while traversing pointer can move in left or in right direction. Use of SkipNet instead of Skiplist provides various advantages like data retrieval, security, improved availability, performance, manageability.

Other overlay networks (Zou et al., 2013) like CAN, CHORD does not provide the control over data i.e. where data is stored and also, they do not provide any guarantee that message routed within same organization remains in that organization. Due to these advantages SkipNet is selected for implementing global indexing.

In this paper, a novel indexing technique based on SkipNet and Octree has been proposed. Key contributions of this study are as follows:

1. A novel indexing technique based on SkipNet and compressed Octree has been proposed. To the best of our knowledge ours is the first work to construct secondary indexing using SkipNet and Octree;
2. Range query algorithms have been proposed;
3. Nearest neighbor query algorithms have been proposed.

Whole paper is organized is as follows: section 2 presents the literature work which has been done so far in indexing field. Section 3 presents the overview of SkipNet-Octree indexing technique. Experiment results and analysis is presented in Section 4.

2. RELATED WORK

This section gives the detailed Current state of work which has been done related to Cloud storage indexing techniques. Google’s Bigtable, Google file system (Ghemawat et al., 2003), Facebook Cassandra (Apache Cassandra, 2013) and many more are the existing scalable storage systems. An open source software Hadoop provides both storage and processing facility for cloud computing. Petabytes, Exabyte of data can be stored on distributed file system with hundreds and thousands of commodity machines. However, Hadoop batch system shows poor performance for range and multidimensional queries. Moreover, current systems support key-value pairs to organize and data retrieval. Key-value pair data retrieval is suitable for exact matching through point queries but not useful for multidimensional, range queries or more complex queries.

In Cloud Storage System, Indexing plays a very important role for efficient query processing. Several indexing techniques have been implemented and proposed (Gani et al., 2016). To organize large scale cloud data, global scalable distributed fault-tolerant B-tree indexing (Aguilera, et al.,
The Theory and Implementation of InputValidator: A Semi-Automated Value-Level Bypass Testing Tool
[www.igi-global.com/article/theory-implementation-inputvalidator/2651?camid=4v1a](www.igi-global.com/article/theory-implementation-inputvalidator/2651?camid=4v1a)

Multiple-User Simultaneous Testing: Experience with Two Methods
[www.igi-global.com/chapter/multiple-user-simultaneous-testing/40507?camid=4v1a](www.igi-global.com/chapter/multiple-user-simultaneous-testing/40507?camid=4v1a)