Expert Systems With Applications 186 (2021) 115713

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Novel distributed load balancing algorithms in cloud storage

Yogesh Gupta [1]

BML Munjal University, India

ARTICLE INFO

Keywords:
Cloud computing
Load balancing
Cloud storage
Distributed system
Cloud services

ABSTRACT

From last decade, there is a rapid expanded upon in the data in cyberspace. In order to manage them efficiently, distributed storage came into the world. Cloud storage is a type of distributed storage based on cloud computing technology. Cloud storage acts as a repository in which data stored, managed and made accessible to users. Largest generated application datasets can flexibly be stored or deleted in the cloud and from here end users access this data by using cloud storage services interface, without accessing any storage server in real. Cloud storage system is considered of hundreds of independent storage servers which are distributed geographically, and handle millions client requests concurrently. Some of the storage servers get huge clients requests and some servers remain under loaded. Due to this unequal distribution of load on storage servers degrades the performance of overall system and increases the response time. This work addresses issues regarding to efficient utilization of storage servers in cloud storage. Handling various challenges regarding to the load balancing in the cloud storage is the one of the main objectives of this research. Though analyzing the contribution of other authors in this area, in this work two distributed load balancing algorithm CDLB and DDLB are proposed by exploiting the different parameter of storage server. The first proposed algorithm considers the service rate and queue length as a main parameter of the server. The second proposed algorithm considers extra server parameter such as service time and deadline time of the client request. This work monitors various aspects which leverage the overall performance of cloud storage. Both the algorithms try to balance the load of storage servers as well as effectively utilize the server capabilities. From the simulation results, it can be concluded that proposed algorithms balance the load, efficiently utilize the server capabilities, reduce the response time and leverage the overall system performance.

1. Introduction

Rapid growth of Internet Technologies has boosted the data proliferation exponentially on the network and cloud storage manages this generated data (Prabavathy et al., 2013). Large generated application datasets can flexibly be stored or deleted in the cloud and from here end users access this data by using cloud storage services interface, without accessing any storage server in real. Cloud storage system is considered of hundreds of independent storage servers which are distributed geographically (You et al., 2011), and millions of client requests are handled by these storage servers concurrently. These storage servers get huge clients requests and some servers remain under loaded. Due to this unequal distribution of load on storage servers leads to degrade the performance of overall system and increases the response time. Therefore, two different load balancing algorithms are proposed in this work to improve the overall system performance.

1.1. Cloud computing

Cloud computing is a new style of web-based computing model for providing flexible, cost-effective and on demand network access to computing capabilities those can be quickly allocated and managed with negligible administrative effort or with least interaction to cloud service provider (Mell & Grance, 2011). It is a pond of manageable, virtual and highly extensible computing infrastructure which enables the customers to host their applications over internet in cost effective manner (James et al., 2008).

Cloud computing is emerging as the most recent disseminated computing paradigm which gives excess, reasonable and adaptable resources on request to client over the internet. This technology effectively exploits the sharable resources on internet such as memory, storage, computation power and bandwidth. Cloud service delivery is divided into three models: Software as a Service (SaaS), Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) (Zhang et al., 2010).

E-mail address: er.yogeshgupta@gmail.com.
[1] 0000-0002-4561-7210

https://doi.org/10.1016/j.eswa.2021.115713
Received 8 March 2021; Received in revised form 6 June 2021; Accepted 1 August 2021
Available online 8 August 2021