International Journal of Scientific Research in Computer Science and Engineering
Vol.6, Issue.3, pp.67-70 , June (2018)
Review Paper
E-ISSN: 2320-7639

Efficient Load Balancing Using Restful Web Services in Cloud Computing: A Review

Tusha Agarwal[1*], Neeta Sharma[2]

[1] Department of Computer Science, Noida International University, Noida, India
[2]School of Engineering and Techonology, Noida International University, Noida, India

**Corresponding Author: aggarwal.tusha.tusha@gmail.com*

Available online at: www.isroset.org

Received: 21/May/2018, Revised: 07/Jun/2018, Accepted: 24/Jun/2018, Online: 30/Jun/ 2018

Abstract- In today's world Internet is the higher source of all kind of information's. Modern high-traffic websites must serve hundreds of thousands request from user to clients and vice versa. These services return the required information in form of text, images, video etc. In Cloud computing, Load Balancing is required in such situations to avoid overload. A load balancer technique mediates client access requests to servers and intelligently decides which server is best placed to fulfil each request. Restful interfaces are mainly used for implementation of web services and are based on the resource-oriented approach. This paper discusses the some existing load balancing algorithms in cloud computing. In this research paper, Restful services are used for data storage and retrieval from Cloud system. Cloud is a storage mechanism in which one can store, process data on demand. Cloud based on service oriented architecture is known as service oriented cloud computing architecture**.** This approach has reduced the amount of data used for recovery to almost half and also maintains a secure access control mechanism for authenticated user.

Keyword- REST, HTTP, Web Services, Load Balance, XOR scheduling

I. INTRODUCTION

Cloud Computing is a new emerging field in the IT environment. It is an Internet-based service that gives access to users to share their resources and any other useful information. Computation in cloud is done with the aim to achieve maximum resource utilization and cost minimization. Cloud computing involves virtualization, distributed computing, utility computing, networking, software and web services.

Cloud storage is built up of numerous inexpensive and unreliable components, which leads to a decrease in the overall mean time between failures. As storage systems grow in scale and are deployed over wider networks, component failures have been more common, and requirements for fault tolerance have been further increased. A cloud system consists of several elements such as clients, data-centre and distributed servers. It has characteristics such as fault tolerance, high availability, scalability, flexibility, reduced overhead for users, reduced cost of ownership, on demand services etc. Effective load balancing algorithm is required in order to cope up with these issues. There can be different types of loads such as CPU load, memory capacity, and delay or network load [8].

Cloud computing is a distributed computing system that focuses on providing a wide range of users with distributed access to scalable, virtualized hardware and software infrastructure over the internet[10]. Cloud computing methodology completely changes the concept of parallel and distributed computing. It provide a very easy solution to all IT resources.This is all suggests that cloud computing will change the way we interact with the resources via Internet. Cloud models use virtualization technology. It depends on the hardware configuration of the data centre or server in how may virtual machine they can be divided. Load balancing is the pre-requirements for increasing the cloud performance and for completely utilizing the resources [9].

Load balancing is the process of distributing the load among various processors to improve resource utilization and the throughput time while also avoiding a situation where some nodes are heavily loaded while other nodes have very less load or are working scarcely. All the processors in the system or every node in the network does approximately equal amount of work at any point of time when load balancing is applied to the system. Load balancing is a pre-required service for increasing the performance and maximum utilization of the resources. Load balancing is the process of increasing system performance in the situations of heavy load. This process of

removing the situation in which some of the nodes are overloaded while some others are under loaded.This phenomenon can drastically reduce the working efficiency [3].

This system performs recovery of data in case of disk failures using Cauchy matrix heuristics. First, it uses Cauchy matrix heuristics to produce a matrix set. Second, for each matrix in this set, it uses XOR schedule heuristics to generate a series of schedules. Finally, it selects the shortest one from all the produced schedules. In such a way, it has the ability to identify an optimal coding scheme, within the capability of the current state of the art, for an arbitrary given redundancy configuration using restful web services.

As a last paragraph of the introduction should provide organization of the paper/article (Rest of the paper is organized as follows, Section I contains the introduction of Cloud computing and load balancing, Section II describes the concept of Restful web services, Section III describes the HTTP protocol , Section IV describes how web services communicate with the web, section V explain the load balancing and load balancing algorithms, Section VI describes XOR scheduling and Caco approach for cloud data and Section VII concludes research work with future directions.

II. REST

Representational state transfer (REST) or Restful web services is one way of providing inter operability between computer systems on the Internet. Restful interfaces are mainly used for web service implementations and are built upon the resource-oriented approach. REST proposes the use of uniform and predefined set of stateless operations to exchange heterogeneous resource representations. REST compliant Web services allow requesting systems to access and manipulate textual representations of Web resources using a uniform and predefined set of stateless operations. The term representational state transfer was introduced and defined in 2000 by Roy Fielding in his doctoral dissertation. Fielding used REST to design HTTP 1.1 and Uniform Resource Identifiers (URI).

The principles of REST include:

1. Conceptual entities and functionalities are modelled as resources identified by universal resource identifiers (URIs).
2. Resources accessed and manipulated via standardized, well-known HTTP operations (GET, POST, PUT and DELETE).
3. Components of the system communicate via these standard interface operations and exchange the representations of these resources (one resource may have multiple representations).

In REST system, servers and clients typically travel through different states of resource representations by following the interlinks between resources.

By applying the principles of REST Web service (WS), development, Restful WSs are emerging as the choice for many of the leading Internet companies to expose their internal data and functionalities as URI identified resources. In contrast to the operation-centric perspective of WSDL/SOAP WSs, Restful WSs view the applications from a resource-centric perspective.

III. HTTP

The Hypertext Transfer Protocol (HTTP) is an application protocol for distributed, collaborative, and hypermedia information systems. HTTP is the foundation of data communication for the World Wide Web. Hypertext is structured text that uses logical links (hyperlinks) between nodes containing text. HTTP is the protocol to exchange or transfer hypertext. HTTP functions as a request response protocol in the client server computing model. A web browser, for example, may be the client and an application running on a computer hosting a website may be the server. The client submits an HTTP request message to the server. The server, which provides resources such as

HTML files and other content, or performs other functions on behalf of the client, returns a response message to the client. The response contains completion status information about the request and may also contain requested content in its message body.

IV. WEB SERVICES

A Web service is a service offered by an electronic device to another electronic device, communicating with each other via the World Wide Web. In a Web service, Web technology such as HTTP, originally designed for human-to-machine communication, is utilized for machine-to-machine communication, more specifically for transferring machine readable file formats such as XML and JSON. In practice, the Web service typically provides an object-oriented Web-based interface to a database server, utilized for example by another Web server, or by a mobile application, that provides a user interface to the end user. Another common application offered to the end user may be a mash-up, where a Web server consumes several Web services at different machines, and compiles the content into one user interface.

V. LOAD BALANCE

A load balancer is a device that distributes network or application traffic across a cluster of servers. Load balancing improves responsiveness and increases availability of applications.

A load balancer sits between the client and the server farm accepting incoming network and application traffic and distributing the traffic across multiple back end servers using various methods. By balancing application requests across multiple servers, a load balancer reduces individual server load and prevents any one application server from becoming a single point of failure, thus improving overall application availability and responsiveness.

Some of the major goals of load balancing algorithms:

a) Cost effectiveness and low energy consumption with improvement in system performance at a reasonable cost.

b) The distributed system in which the algorithm is implemented may change in size or topology. Hence, scalable and flexible algorithms should be used to allow such changes to be handled easily. But load balancing is critical to serve requests cost effectively over cloud. So, to overcome limitations of balancing algorithm, we propose to implement re-balancing algorithm. Before serving requests over cloud, it will calculate performance and load on individual resources. Based on this calculated performance and current load, requests will be served by those specific resources over cloud.

Different load balancing algorithms provide different benefits; the choice of load balancing method depends on your needs:

Round Robin–Requests are distributed across the group of servers sequentially.

Least Connections–A new request is sent to the server with the fewest current connections to clients. The relative computing capacity of each server is factored into determining which one has the least connections.

IP Hash–The IP address of the client is used to determine which server receives the request.

Algorithm1: Load Balancer Algorithm
Input: Text files with data
Output: File operation with server load balancing
Step 1: Initialize server and its sub-servers
Step 2: Establish connection between sub-server and servers using the IP or Port number.
Step 3: Upload File to server that should be shared.
Step 4: Server encrypts data with MD5 Encryption.
Step 5: Split the file into multiple chunks
Step 6: Calculate each sub server memory
Step 7: Divide the total chunks value by total number of sub-servers
Step 8: Upload each chunk into sub servers based on its memory capacity
Step 9: If Capacity is less then transfer the excess chunks into next sub-servers
Step 10: Each chunk will be appended with an index value.
Step 11: When the client request for a file, that will be received from different sub-servers based on the index value.
Step 12: Client collects all the chunks then the file will be decrypted, then that will be viewed by client.

Algorithm2: Equally Load Re-Balancer Algorithm
Input: Each Node load base on current hitting
Output: Distributed data to each node.
Step 1: Initialize all data nodes which are connected to master node as n.
Step 2: for each (1 up to n)
Take each ith node server load.
A[i] => ith Node load degree or hitting load.
End for.
Step 3: get total length of A. create the data requested chunks.
K=A. length ();
Step 4: Generate k mappers for distribute a data.
Step 5: Assign each chunks to each mapper.
Step 6: Request to server for saving data.
Step 7: end procedure.

Algorithm3: Caco Matrix Generation
Input: Data block d
Output: Cauchy matrix as X
Step 1: Constructing the matrix ONES. First, CaCo constructs a matrix named ONES, whose element (i,j) is defined as the number of ones contained in the binary matrix M(1/i+j).
Step 2: Choosing the minimal element. Second, CaCo chooses the minimal element from the matrix ONES. Supposing the element is (x1,y1), we initialize X to be {x1} and Y to be Y1.
Step 3: Determining the set Y. Besides the element (x1,y1) CaCo chooses the top k-1 minimums from row x1. Then, CaCo adds the corresponding k-1 column numbers to Y , and we have Y={y1,y2,y3....yn}.
Step 4: Finally generate matrix as X for each row each X instance is Cauchy matrix of given data block.

VI. XOR SCHEDULING

As the amount of data increases exponentially in large data storage systems, it is crucial to protect data from loss when storage devices fail to work. Recently, both academic and industrial storage systems have addressed this issue by relying on erasure codes to tolerate component failures.

The traditional XOR-scheduling algorithm follows the intuitive idea that coding words should be produced one by one. Instead, we can reorder the schedule so that it consumes data words one by one. Our new XOR-scheduling algorithm is based on this idea, and its characteristics are as follows:

1. The order of XOR operations is guided by the order of data words instead of coding words.
2. Each data word is used for all of its coding calculations before moving onto the next data word in the same packet.

Caco Approach for cloud data :-

The CaCo model is used for cloud data hierarchy analysis and monitoring. The system consist of a web service server which performs all the communication. All the requests to the databases are routed through web service server. Client requests are sent to the server which redirects them to the appropriate node. CaCo matrix is stored in data node and provides data in case of node failure. The system proposed in this architecture consists of a semi-automated stack for data movements and address tracking. This system majorly focuses on data waiting time reduction and performance enhancement.

Step 1: Constructing the matrix ONES. First, CaCo constructs a matrix named ONES, whose element (i,j) is defined as the number of ones contained in the binary matrix M(1/i+j).

Step 2: Choosing the minimal element. Second, CaCo chooses the minimal element from the matrix ONES. Supposing the element is (x1,y1), we initialize X to be x1 and Y to be Y1.

Step 3: Determining the set Y. Besides the element (x1,y1) CaCo chooses the top k-1 minimums from row x1. Then, CaCo adds the corresponding k-1 column numbers to Y , and we have Y=y1,y2,y3....yn.

Step 4: Finally generate matrix as X for each row each X instance is cauchy matrix of given data block.

VII. CONCLUSION

Cloud computing is now becoming a business standard. It simplies the users accessibility. It provides a virtual storage space to the user which could be used without bothering about the details of the entire mechanism. In this review paper, restful web services are used for communication and system focus on CaCo, an approach that incorporates all existing matrix and schedule heuristics, and thus is able to identify an optimal coding scheme within the capability of the current state of the art for a given redundancy configuration. The selection process of CaCo has an acceptable complexity and can be accelerated by parallel computing. It should also be noticed that the selection process is once for all.

For the future enhancement we can consider as load rebalancing in hybrid cloud environment with Hadoop. The thermal management and energy saving approach with resource virtualization is another interesting area for such concepts. The paper also explain that use of restful web services for all communications have increased the performance by reducing the data transfer time, load balancing has made the efficient use of processors.

REFERENCES

[1] Zhang, G., Wu, G., Wang, S., Shu, J., Zheng, W. and Li, K ,"An Efficient Cauchy Coding Approach for Cloud Storage Systems". *IEEE Transactions on Computers* ,vol. 65, no.2, pp.435-447, 2016.

[2] Trifonov, P, " Low-Complexity Implementation of RAID Based on Reed-Solomon Codes", *ACM Transactions on Storage*, vol. 11, no.1, pp.1-25, 2015.

[3] Li, X., Zheng, Q , Qian, H., Zheng, D. and Li, J, "Toward optimizing cauchy matrix for cauchy reed-solomon code", *IEEE Communications Letters*, vol.13, no.8, pp.603-605, 2009.

[4] Gruner, S., Pfrommer, J. and Palm, F, "RESTful Industrial Communication With OPC UA", *IEEE Transactions on Industrial Informatics*,vol.12, no.5, pp.1832-1841,2016.

[5] Luo, C, Zheng, Z., Wu, X., Yang, F. and Zhao, Y., " Automated structural semantic annotation for RESTful services", *International Journal of Web and Grid Services*, vol.12, no.1, p.26,2016.

[6] Zhao, X., Liu, E., Yu, H. and Clapworthy, G. "A linear logic approach to the composition of RESTful web services", *International Journal of Web Engineering and Technology*, vol.10, no.3, p.245,2015.

[7] M. Sharma, A. Yadav, P. Sharma, "*An Optimistic Approach for Load Balancing in Cloud Computing*", International Journal of Computer Sciences and Engineering, Vol.2, Issue.3, pp.26-30, 2014.

[8] Microsoft Windows Azure, "Developing Applications for Highly Available Storage of Cloud Service", *International Journal of Science and Research (IJSR)*, vol. 4, no.12, pp.662-665.

[9] M. Lagwal, N. Bhardwaj, "*A Survey On Load Balancing Methods and Algorithms in Cloud Computing*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.4, pp.46-51, 2017.

[10] R.S Sajjan, R.Y. Biradar, "*Load Balancing and its Algorithms in Cloud Computing: A Survey*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.1, pp.95-100, 2017.

