

Spatio Temporal Load Re-balancing Approach in Cloud Computing Applications

Sheetal K. Gajbhiye, Prof. Rugraj Purohit

Department of Computer Engineering, Alard College of Engineering and Management, Pune, India

ABSTRACT: Distributed computing is fueled by a motor known as Internet Data Centers (IDC). As distributed computing prospers, the vitality utilization and expense for IDCs are taking off. The vitality cost minimization issue for IDCs in deregulated power markets has produced developing premium. In this paper we concentrate how to influence both the geographic and fleeting variety of vitality cost to minimize vitality cost for dispersed IDCs. We propose a novel structural planning and two calculations for bound together spatial and fleeting burden adjusting. Thorough examination demonstrates that our calculations have a low computational many-sided quality, require a casual exactness in power value estimation, and assurance an administration culmination time for client demands. Utilizing genuine power cost and workload follows, broad assessments exhibit that contrasted with the plans utilizing either spatial burden adjusting or worldly load adjusting alone, the proposed spatio-fleeting burden adjusting strategy essentially decreases vitality cost for conveyed IDCs. For future work, we have recognized bottlenecks in Load Balancing Algorithm and proposed to execute Load Re-Balancing Algorithm. Re-Balancing Algorithm will compute Server Performance and taking into account that it will re-equalization work demands. This will ensure there is change accordingly time for administration demands.

KEYWORDS: Data center, energy, cost management, delay bound, load balance, algorithm.

I. INTRODUCTION

Big data [9] is the term for a gathering of data sets so huge and complex that it becomes hard to process using on hand database management tools or traditional data processing Environments. The difficulties incorporate storage, search, sharing, transfer, analysis, and visualization, exchange, examination, and perception. The pattern to bigger information sets is because of the extra data logical from investigation of a solitary extensive arrangement of related information, when contrasted with particular littler sets with the same aggregate sum of information, permitting relationships to be found to "spot business patterns, decide nature of exploration, forestall maladies, join lawful references, battle wrongdoing, and decide constant roadway movement conditions. The confinements additionally influence Internet inquiry, fund and business informatics. Information sets develop in size partially in light of the fact that they are progressively being assembled by omnipresent data detecting versatile ethereal sensor advances (remote detecting), programming logs, cameras, mouthpieces, ID perusers, and remote sensor systems. Huge information is hard to work with utilizing most social database administration system and desktop measurements and perception bundles, requiring rather "greatly parallel programming running on a n number of servers". What is viewed as "large information" [14] changes relying upon the abilities of the association dealing with the set, and on the capacities of the applications that are customarily used to prepare and examine the information set in its area. "For a few associations, confronting many gigabytes of information interestingly might trigger a need to rethink information administration alternatives. For others, it might take tens or many terabytes before information size turns into a huge thought [10].

II. LITRETURE SURVEY

Cloud computing is an effective way for managing, maintaining and analyzing huge amount of data. Lingjuan Li et al [1] learned multiple parallel association rule mining approach. Apriori base approach was found to be wasting much memory space and time while handling large data. So improved Apriori algorithm [1] was proposed which partitioned

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

the data into chunks and applied association mining on each chunk in parallel. The results were then integrated for identifying global frequent item set. This algorithm was developed and tested on Hadoop [1] platform with parallel processing [1] programming model and its performance was reviewed.

Mr. Vipul Agarwal et al [2] dealing with some of the issues in association rule mining like security and privacy. The primary aim of privacy preservation is to extract exact patterns without accessing original data. For modification purpose, the item / the chunk is selected from the chunk set which would have minimal influence on original database. For continuing privacy of users a new algorithm named Cloud Service Preserving Model algorithm is proposed in this paper. This algorithm authorizes the correct user and hides his logical instances from others. The algorithm selection depends on situation for reducing the influence for original database. The proposed algorithm is useful in multiuser cloud computing environment.

Multi-Agent Systems (MAS) are variant systems with multiple agents looking for a solution to a common problem. The main goal in this paper is to achieve an efficient data mining process with an extensible and general architecture which can maintain multiple data types. Othmane et al [3] showed a multi-agent architecture for cloud computing as Software as a Service (SaaS) [3]. Major issues faced in achieving a scattered mining procedure with beyond mentioned characteristics were enormous data volumes, communication issues, heterogeneous nature of users and complete parallelism. Those challenges are overcome by multi-agent approach. The process of classification originate rules for categorizing records. H. Malmiret al [4] presented a difference in between two classification algorithms in conditions of speed and correctness using Genetic Algorithm (GA) classifiers. Imperialist Competitive Algorithm (ICA)[4] and Particle Swarm Optimization (PSO)[4] both the algorithms are used for improving classification speed and accuracy. Multi-Layer Perception (MLP) [4] is the neural network that was used for analyzing and evaluating these algorithms. data which is divided in two parts and processing these part on being systems which were interconnected due to this Parallel processing was achieved .

Z. Hai-Jian et al [5] examined shopper highlight extraction with distributed mining so as to compute their inclinations. Distributed computing gives monstrous capacity limit and adaptable registering ability. Procurement of exact administrations with proper innovation is the way to accomplishment for ventures when managing enormous information. Information mining in distributed computing is an effective and solid strategy for removing helpful examples from huge information. Parallel preparing decreases the undertaking's handling cost as it is not reliant on execution of machines. Parallelism enhances handling speed and gives adaptation to non-critical failure in huge information mining. In any case, this methodology has a few difficulties, for example, calculation determination, instability in calculation and assessment of results. By knowing the inclinations of clients, endeavor can reshape their association with client however it represents a genuine risk to security and protection with regards to the individual data of customers.

III. PROBLEM STATEMENT

We model the dispatching and execution of user requests for cloud-computing service. First we describe portal server, and distributed IDC system. Next we explain workload of user requests, electricity price, workload queue, and workload scheduler. Then we give the queuing delay constraint for workload and the capacity constraint for IDC sites. We further formulate the energy cost minimization problem for the distributed IDC system

IV. PROPOSED SYSTEM

In this structural engineering a brought together load balancer is utilized to part the record into pieces keeping in mind the end goal to store the information in different hubs. The information to be put away in the cloud is scrambled before capacity for more security. The encryption is finished by the key created at the customer side. At that point the scrambled information is made into lumps and put away in different hubs. At the point when the server control performs operations on information such as cancellation or updating load awkwardness issue happens. This issue can be explained by the rebalancing calculation which adjusts the heap in the cloud after the above operations performed. As the unified hub has the reinforcement of the erased documents it can store the erased documents in alternate hubs even for the situation of hub disappointment.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

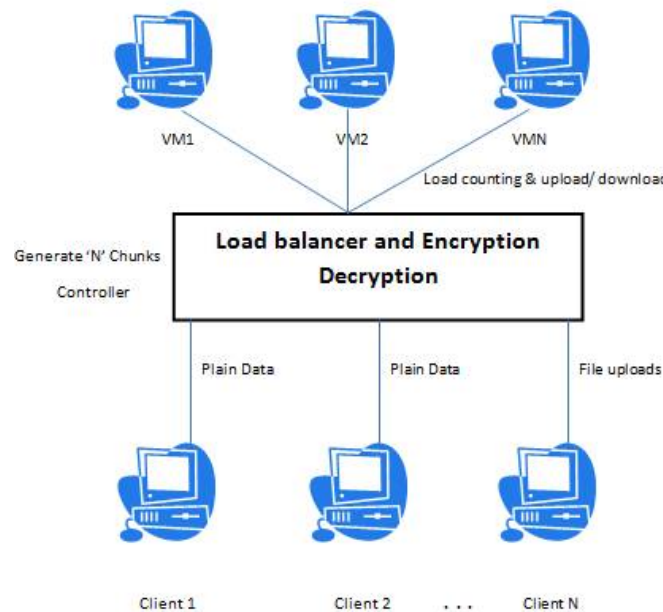


Figure 1: Proposed System Architecture

The data that is to be stored in cloud is not protected. In order to afford security to data, the data is stored in the encrypted shape in the nodes. The file that is to be uploaded in the cloud is chosen by the client. The encryption procedure is performed over the data with the key provided. The encrypted file is finished in different chunks and stored in different nodes.

ALGORITHMS

Compute VM Load from data nodes

Input: i th Node input

Output: - Idle or Normal Or Overloaded in percent (%).

Compute Load (VM id) : weight Degree Inputs: The static parameter comprise the number of CPU's, the CPU dispensation speeds, the reminiscence size, etc. active parameters are the memory consumption ratio, the CPU exploitation ratio, the network bandwidth. procedure:-

1. Characterize a load limit set: $F = \{F_1, F_2 \dots F_m\}$ with each Fire present the total number of the consideration.

2. Calculate the load capacity as weight Degree $(N) = \sum a_i F_i$ Where $i = 1 \dots m$

3. ordinary cloud partition degree from the node consignment degree statistics as:

$$\text{Load amount avg} = \sum_{i=1}^n$$

LoadingDegree(N_i)

4. Three height node position are defined Load_degree(N)=0 for inactive $0 < \text{Load_Degree}(N) \leq \text{Load_Degree}(N)$ for overfull

Equally Spread Current Execution throttled Load balancing algorithm:

Input : File form user as F_i

Output: Equally distributed chunks on data servers

Step1 : Read F_i from data owner with size

Step 2: count total number of data nodes N_i

Step 3: for each(score=read each vm node and call to compute node(k))

Read when $k == \text{null}$

End for

Step 4: create data chunks base on server loads score.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Step 5: save all data on data nodes.

MATHEMATICAL MODEL

Inputs:

The static parameters include the number of CPU's, the CPU processing speeds, the memory size, etc. Dynamic parameters are the memory utilization ratio, the CPU utilization ratio, the network bandwidth.

Process:-

1. Define a load parameter set: $F = \{F_1, F_2, \dots, F_m\}$ with each F_i presents the total number of the parameters.

2. Compute the load degree as

$$\text{Load Degree}(N) = \sum_{i=1}^m a_i F_i$$

Where $i = 1 \dots m$

3. Average cloud partition degree from the node load degree statistics as:

$$\text{Load degree}_{avg} = \sum_{i=1}^n \text{Load Degree}(N_i)$$

4. Three level node status are defined

Load degree(N)=0 for **Idle**

$0 < \text{Load Degree}(N) < \text{Load Degree}(N)_{high}$ for **Normal**

$\text{Load Degree}(N)_{high} \leq \text{Load Degree}(N)$ for **Overloaded**

Output :-

Idle or Normal Or Overloaded

Results and Discussion

The system is work with multiple virtual machines. One we deploy the system on EC2 environment it will work fine. The estimated results for the system base on below tables.

VM Name	CPU Utilization	Chunk Size	
		2.8 MB	5.6 MB
VM1	50%	2.8 MB	5.6 MB
VM2	60%	2.8 MB	5.6 MB
VM3	60%	2.5 MB	5.0 MB
VM4	75%	1.4 MB	2.8 MB
VM5	85%	0.8 MB	1.6 MB
	Total Size	10.0 MB	20.0 MB

Table of system performance

Below graph show the system performance of proposed system retrieval time for total number of files from data servers.

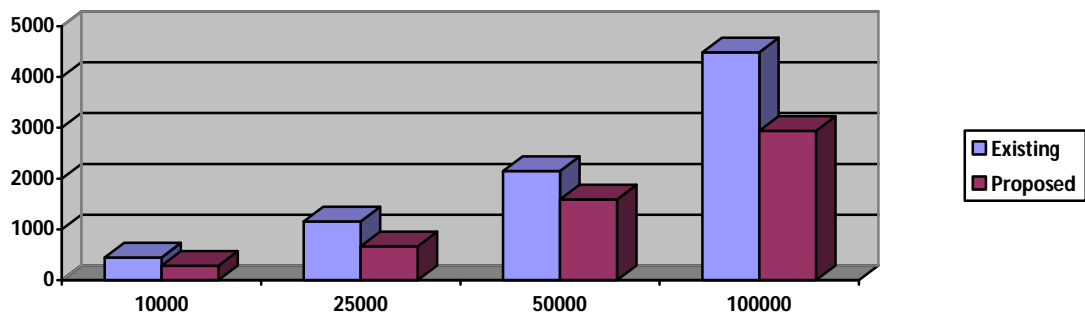


Fig 2: Time required between proposed vs existing

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

V. CONCLUSION

Current Internet data centers consume enormous amount of energy and incur huge energy cost Load Balancing helps to reduce cost of operation of IDC by effectively distribute load based on Space and Time parameters. We have proposed to implement Load Re-Balancing Algorithm which will further improve performance, time, space and will help in reducing energy cost of IDC.

Future Scope

Load balancing is a idea which is still beneath research. on a daily basis new frameworks, algorithms and models are creature developed and existing models are updated. There is a huge scope for future improvement. For example, the users are distribution illogical data as a query on the web, and therefore web-centric queries can be optimized at server side to decrease server load to develop the server performance. Further, implementation of our work is awaiting and hence an improvement may be recommended in the identical.

REFERENCES

- [1]. L. Li and M. Zhang, 'The Strategy of Mining Association Rule Based on Cloud Computing', 2011 International Conference on Business Computing and Global Information, 2011.
- [2]. V. Agarwal, Y. Khandagre, and A. K. Dubey, 'Novel Cloud Subset Preserving Mining (CSPM) algorithm for association rule mining in centralized database', 2012 IEEE International Conference on Technology Enhanced Education (ICTEE), 2012.
- [3]. Othmane, B., R. Hebri, and M. Boudiaf, 'Cloud Computing and Multi-Agent Systems: A new Promising Approach for Distributed Data Mining', Proc. of the ITI 2012 34th Int. Conf. on Information Technology Interfaces, Jun 25-28, 2012.
- [4]. H. Malmir, F. Farokhi, and R. Sabbaghi-Nadooshan, 'Optimization of data mining with evolutionary algorithms for cloud computing application', ICCKE, 2013.
- [5]. Z. Hai-Jian, 'Analysis and Application of Consumer Features with Cloud Computing and Data Mining Technology', 7th International Conference on Intelligent Computation Technology and Automation, 2014.
- [6] T.R. Gopalakrishnan Nair et al, "Data Mining using Hierarchical Virtual K-means approach integrating data fragments in Cloud Computing Environment" PROCEEDINGS OF IEEE CCIS, 2011
- [7] Qiwei Lu et al, "Secure Collaborative Outsourced Data Mining with Multi-owner in Cloud Computing" IEEE 11TH INTERNATIONAL CONFERENCE ON TRUST, SECURITY AND PRIVACY IN COMPUTING AND COMMUNICATIONS, 2012
- [8] JianlinXu et al, "MobSafe: Cloud Computing Based Forensic Analysis for Massive Mobile Applications Using Data Mining" TSINGHUA SCIENCE AND TECHNOLOGY, Volume 18, Number 4, August 2013
- [9] Yong Wang et al, "Transplantation of Data Mining Algorithms to Cloud Computing Platform when Dealing Big Data", INTERNATIONAL CONFERENCE ON CYBER-ENABLED DISTRIBUTED COMPUTING AND KNOWLEDGE DISCOVERY, 2014
- [10] Xin Song et al, "An Integrated Framework for Analysis and Mining of the Massive Sensor Data Using Feature Preserving Strategy on Cloud Computing" SEVENTH INTERNATIONAL SYMPOSIUM ON COMPUTATIONAL INTELLIGENCE AND DESIGN, 2014.