Volume 5, Issue 10, October-2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Effective Data Centre Management Policy for Cost Reduction

K. Jayanth
PG Student
Department of Computer Science and Engineering
Madanapalle Institute of Technology & Science
Madanapalle, Andhra Pradesh, India

M. Veeresh Babu
Assistant Professor
Department of Computer Science and Engineering
Madanapalle Institute of Technology & Science
Madanapalle, Andhra Pradesh, India

Abstract: ***The intense growth of requirements for big data systems enforces more pressure on storage, calculations and interaction in data centers, which then happens upon significant functional expenses to data center providers. So, fetched lessening is one of the exploration issues for the era of the huge data period. Dissimilar to routine distributed computing, one of the primary sorts of enormous data administrations is the restricted joining between data gathering and counts as estimation ventures can be performed just when the steady data is accessible. From the outcome, there we discover three angles, i.e., process task, data positioning and data movement, extraordinarily influence the utility costs of data center. The model is going to study the cost minimization problem in the above mentioned three aspects of big data in Geo-distributed data centers. The proposed model masks the solution by VM placement approaches that consider position of data centers and VM migration expenses while improving VM positions over a little set of data centers. This strategy is source intense and as the number of VMs compared to actual hosts improves, the solution time also improves and is very useful in circumstances where real-time scheduling choices are required. To decrease the solution time we recast the problem into Round Robin Algorithm and Throttled Algorithm and Compare the solution time.***

Keywords- cloud computing, virtual machines (VMs), process task, data positioning, data movement, Round Robin Algorithm, Throttled Algorithm

I. INTRODUCTION

Big information is a one amongst the rising hot analysis topic as a result of it's largely utilized in information center application in human society, like government, climate, finance, and science. Currently, most analysis work on massive information falls in data processing, machine learning, and information analysis. The name itself contains the which means of information are thus massive in large volume of each structured and unstructured data gift. The challenges embody capture, duration, storage, search, sharing, transfer, analysis and visualisation [1] Currently large information available as a result of the extra data derived from analysis of massive group of connected information, like contrasted on various tiny groups with constant total quantity of knowledge, allowing parallel relation to be found to identify commercial trends, stop diseases, combat crime therefore on [2]. It is very hard to practice maximum no of electronic information service management systems and desktop statistics and image packages, requiring instead of "tens, hundreds, or maybe thousands of servers are massively parallel code executing". the capabilities of the applications and their unit of measurement traditionally used to analyze the knowledge group in its domain [3]. Big info may be a moving target; what is thought of to be "Big" currently will not be therefore years ahead. "There is a few organizations, facing several thousand of megabytes knowledge of information it is the time to rethink data management decisions. For others, it
ought to take tens or several thousands of gigabytes before info size becomes an enormous thought. Every day, 2.5 large integer bytes of information ar created and ninety p.c of the info within the world these days were made among the past 2 years. Another illustration is our capacity for data era has ne'er been hence intense and substantial since the time that the development of the learning innovation inside of the mid nineteenth century. Case for expansive data, on four Gregorian schedule month 2012, the essential presidential dialog between President Barack Obama and Governor Mitt Romney activated more than ten million tweets among two hours.[4] There is still a positive crevice in execution with relative databases. rising the execution of Map empty and improving the measure of your time nature of huge scale procedure have gotten a goliath measure of consideration, with Map flatten parallel writing computer programs being connected to numerous machine learning and handling calculations. Preparing calculations as a rule should be constrained to over the training data for acquiring the insights to determine or streamline model parameters. It requires serious figuring to get to the substantial scale information regularly.[5]To enhance the proficiency of calculations. Geo-appropriated server farms are worked by numerous associations, for example, Google and Amazon are the powerhouses behind numerous Internet-scale administrations. They are conveyed over the Internet to give better inactivity and excess.

II. RELATED WORK

A. SERVER COST MINIMIZATION

Large-scale knowledge centers are deployed everywhere the World providing services too lakhs of users.As per the[6], knowledge center might encompass massive numbers of servers; those servers are work by using millions of watts of

power. Millions dollars on electricity have to exhibit a significant burden on the operating cost for knowledge center suppliers. Then they are mainly focused to reduce the cost of electricity in data center maintenance [5], [7][8].the functionalities that are point to knowledge center energy management are mainly focused on the techniques called task placement and data center resize(DCR).Data center resize and task placement sometimes together thought-about to match the computing demand. [4] Liu et al. Re-examine by taking network delay into consideration. Fan et al. [10] revise power provisioning methods

On what proportion computing instrumentality are often safely and efficiently hosted at intervals a given power financial resources. Rao et al. [3] explore the best approach to scale back power cost by directing client solicitations to geo-circulated information focuses with subsequently redesigned sizes that match the solicitations. Recently, Gao et al. [11] propose the ideal workload administration and taking so as to level record of inertness, vitality utilization and power costs. Liu et al. [12] scale back power cost and natural Impact utilizing an all encompassing methodology of livelihood levelling that Integrates renewable gives dynamic assessment, and cooling supply.

B. BIG DATA MANAGEMENT

The main concern in huge information managing is dependable and effective information placement. In the direction of this goal, Sathiamoorthy et al.[8] gift a unique family of erasure codes that area unit with efficiency fixable and supply higher dependableness compared to Reed-Solomon codes. They additionally associate to show that their codes area unit optimum on an known trade-off between neighbourhood and minimum distance. Yazd et al. [9] assemble utilization of adaptability inside of the data piece situation strategy to develop vitality strength in data focuses and propose an arranging algorithmic principle.

C. DATA PLACEMENT

study a way to confirm a placement of Video-on-Demand (VoD) copies on the servers and the measure of burden ability distributed to each duplicate in this way as to minimize the correspondence value though making certain the client aptitude. Cidon et al. [13] invent Min Copy sets, a data replication placement theme that decouples knowledge distribution and replication to boost the info sturdiness properties in distributed knowledge centers. Jin et al. [10] recommend a joint advancement topic that in the meantime streamlines virtual machine (VM) arrangement and system stream directing to expand vitality reserve funds.

Algorithm:

1. Round Robin

Step1. Round Robin VM load Balancer maintains an index of VMs. At start all VM's have zero allocation.

Step2.

a. The data center controller receives the user requests/cloudlets.

b. The requests are allocated to VMs in circular way.

c. The round robin VM load balancer will allocate the time quantum for user request execution.

Step3. After the execution of cloudlets, the VMs are de- allocated by the Round Robin VM Load balancer.

Step4. The data center controller checks for new /pending/waiting requests in queue.

Step5. Continue from step-2.

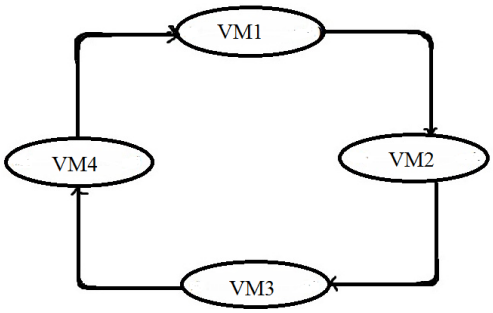


Fig 2.1 positioning virtual machines using RR

2. Throttle Load Balancing:

Step1. Throttled VM Load Balancer keeps an index, table of VMs and the state of the VM (BUSY/AVAILABLE). At the beginning all VM's are available.

Step2. Data Center Controller receives a new request.

Step3. Data Center Controller queries the Throttled VM Load Balancer for the next allocation.

Step4. Throttled VM Load Balancer parses the allocation table from the top until the first available VM is found or the table is parsed completely.

If found:

i) The Throttled VM Load Balancer returns the VM id to the Data Center Controller.

ii) The Data Center Controller sends the request to the VM identified by that id.

iii) Data Center Controller notifies the Throttled VM Load Balancer of the new allocation.

iv) Throttled VM Load Balancer updates the allocation table accordingly.

If not found:
i) The Throttled VM Load Balancer returns -1.
ii) The Data Center Controller queues the request.
Step5. When the VM finishes processing the request, and the Data Center Controller obtains the response cloudlet, it informs the Throttled VM Load Balancer of the VM de-allocation.
Step6. The Data Center Controller checks if there are any waiting requests in the queue. If there are, it carries on from step 3.
Step7. Continue from step 2

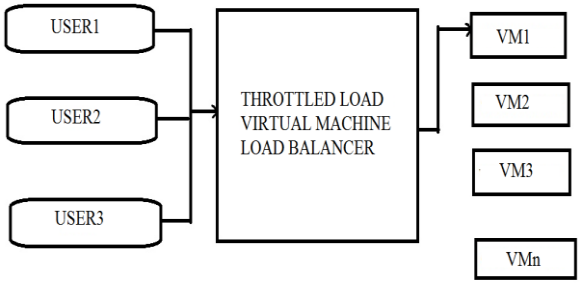
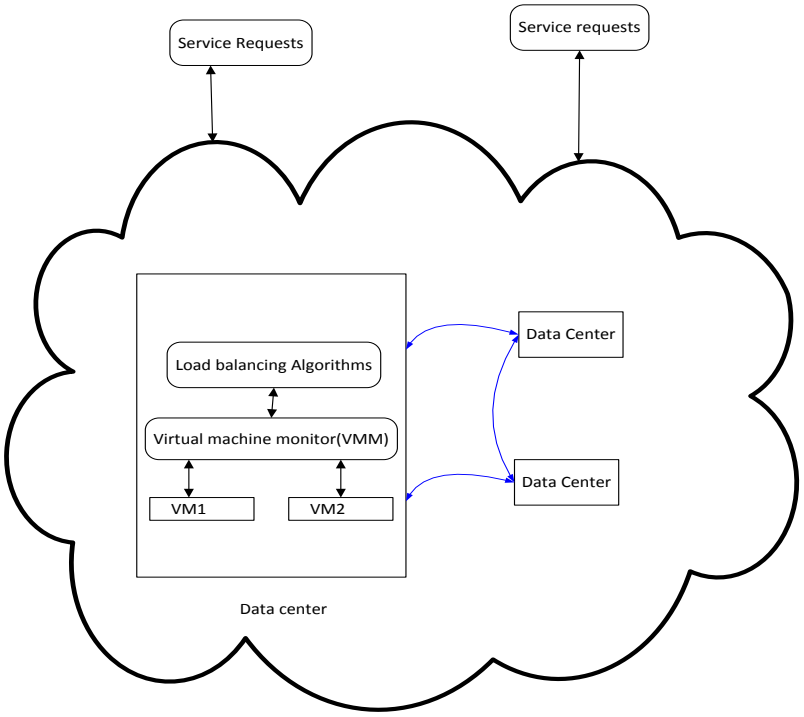


Fig 2.2 Throttled virtual machine load balancer

III. SYSTEM ARCHITECTURE



IV. EXPERIMENTAL RESULTS

4.1. HOME PAGE SCREEN

The following screen shot is showing home page of cloud analyst it is also called as cloud simulator

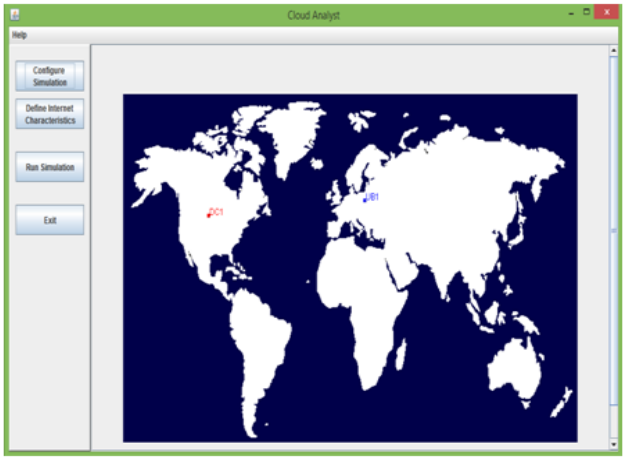


Fig:4.1 Home page

4.2 Loading the Parameters

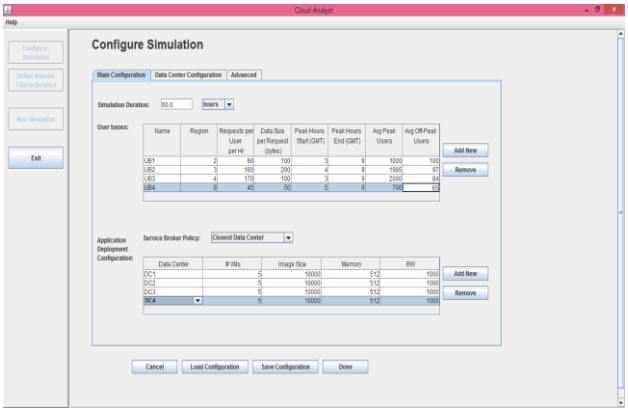


Fig 4.2 UB&DC parameters are to be create

4.3 DC Parameters

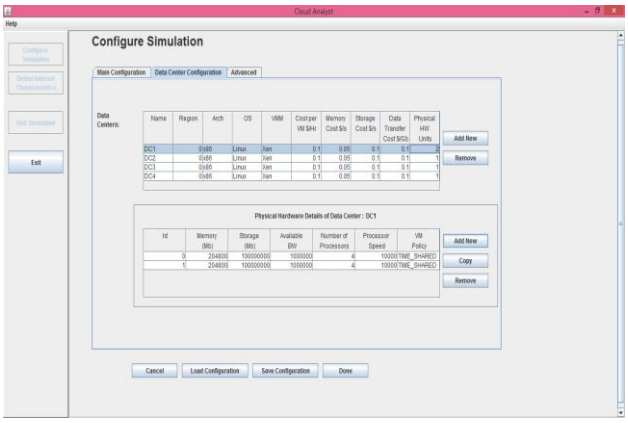


Fig 4.3 DC parameters are to be given as per our requirement

4.4 Loading the Balancing Policy Algorithm into Environment

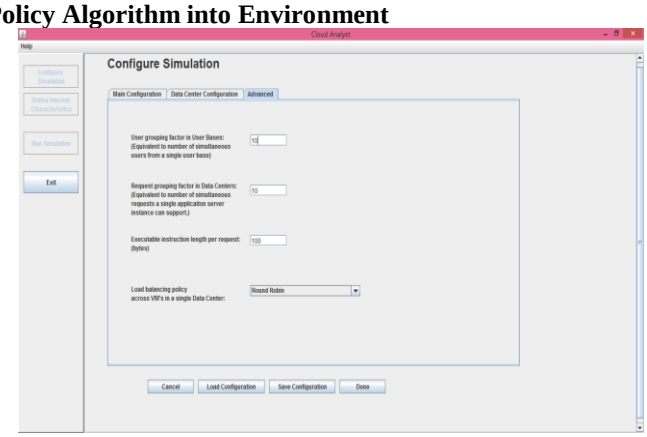


Fig 4.4 Loading algortithm into cloud Analyst

4.5 Response Time of Each Data Center

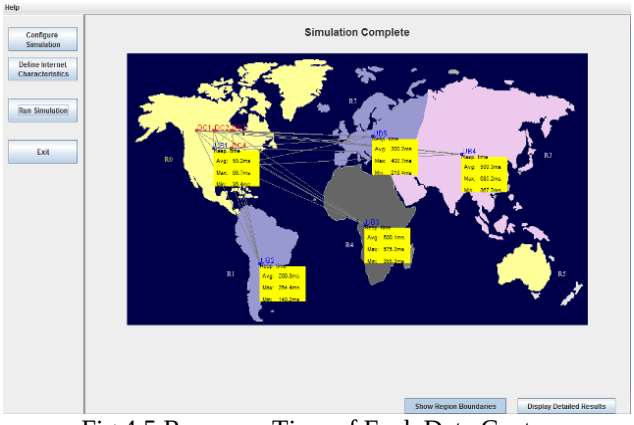
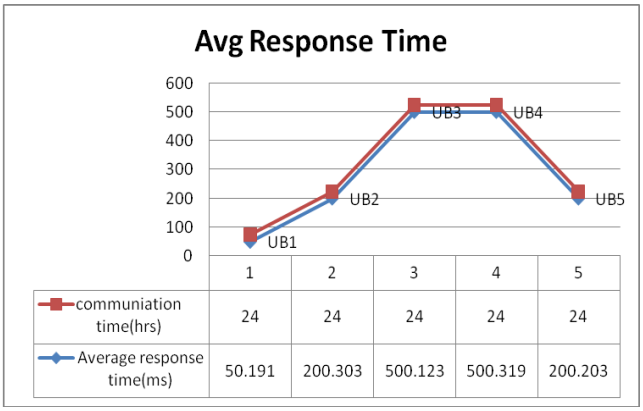


Fig 4.5 Response Time of Each Data Center

If user want to down load file then he has to select the server as above.

4.6 Overall Response Time

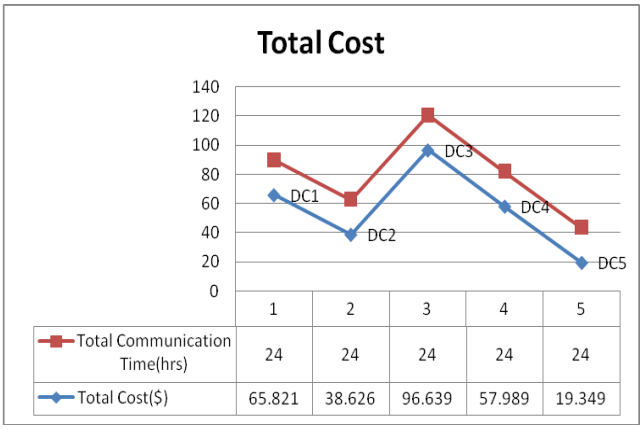


	1	2	3	4	5
communiation time(hrs)	24	24	24	24	24
Average response time(ms)	50.191	200.303	500.123	500.319	200.203

Fig: 4.6 Overall Response.

UB-user base; **ms**- milli seconds; **hrs**- hours

4.7 Overall Cost of Data centers



	1	2	3	4	5
Total Communication Time(hrs)	24	24	24	24	24
Total Cost($)	65.821	38.626	96.639	57.989	19.349

Fig 4.7 Total Cost of Data Center

DC-data center; hrs-hours; $-Dollars

V. CONCLUSION

The proposed model masks the solution by VM placement approaches that consider position of data centers and VM migration expenses while improving VM positions over a little set of data centers. This strategy is source intense and as the number of VMs compared to actual hosts improves, the solution time also minimized and is very useful in circumstances where real-time scheduling choices are required. To decrease the solution time we recast the problem into Round Robin Algorithm and Throttled Algorithm and Compare the solution time. The algorithms performed well in reducing the communication cost and managing the resources over the data center. The data center placement is effectively managed by the load balancing mechanism. The round robin algorithm manages to decrease the overall cost for data centers when compared with the throttled algorithm.

VI. FUTURE ENHANCEMENT

In future, we consider the capacity of machine which require larger than available resources .So, we want to estimate the arrival machine capacities before going to allocation of resources.

REFERENCES

[1] Deelman, E. and Singh, G. and Livny, M. and Berriman, B. and Good, J The Cost of doing Science on the Cloud: The Montage Example, International Conference for High Performance Computing, Networking, Storage and Analysis(SC), 1-12 (2008)

[2] Kondo, D. and Javadi, B. and Malecot, P. and Cappello, F. and Anderson, D.P., Cost-bene_t analysis of Cloud Computing versus Desktop Grids , IEEE International Symposium on Parallel and Distributed Processing(IPDPS), 1-12 (2009)

[3] http://hadoop.apache.org/

[4] J. Dean and S.Ghemawat, MapReduce:simpli_ed data processing on large clusters, Communications of the ACM, 107-113 (2008)

[5] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal Power Cost Management Using Stored Energy in Data Centers," in Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS). ACM, 2011, pp. 221–232.

[6] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the Electric Bill for Internet-scale Systems," in Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM). ACM, 2009, pp. 123–134.

[7] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar, "Benefits and Limitations of Tapping Into Stored Energy for Datacenters," in Proceedings of the 38th Annual International Symposium on Computer Architecture (ISCA). ACM, 2011, pp. 341–352[18] A. Cidon, R. Stutsman, S. Rumble, S. Katti, J. Ousterhout, and M. Rosen-blum, ``Min Copysets: Derandomizing replication in cloud storage," in *Proc. 10th USENIX Symp. NSDI*, 2013, pp. 1_5.

[8] M. Sathiamoorthy *et al.*, ``Xoring elephants: Novel erasure codes for big data," in *Proc. 39th Int. Conf. Very Large Data Bases*, 2013, pp. 325_336.

[9] S. A. Yazd, S. Venkatesan, and N. Mittal, "Boosting energy efficiency with mirrored data block replication policy and energy scheduler," SIGOPS Oper. Syst. Rev., vol. 47, no. 2, pp. 33–40, 2013.

[10] X. Fan, W.-D. Weber, and L. A. Barroso, ``Power provisioning for a warehouse-sized computer," in *Proc. 34th Annu. ISCA*, 2007,pp. 13_23.

[11] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav, ``It's not easy being Green," in *Proc. ACM SIGCOMM*, 2012, pp. 211_222.

[12] Z. Liu *et al.*, ``Renewable and cooling aware workload management for sustainable data centers," in *Proc. Int. Conf. Meas. Model. Comput. Syst.*,2012, pp. 175_186.

