ENERGY EFFICIENT AND REDUCTION OF POWER COST IN GEOGRAPHICALLY DISTRIBUTED DATA CARDS

M Vishnu Kumar[1], E Vanitha[2]

[1] PG Student, [2] Assistant Professor, Department of Computer Science and Engineering,
PTR College of Engineering and Technology, Tamilnadu, (India)

ABSTRACT

Data centers is the most discussed topic in current technical world which considered the brain of a company that holds the responsibilities of storage, management and dissemination of data etc. In other words global technology cannot be feasible without a proper functioning of those data centers. We consider a stochastic optimized method for the process of job scheduling and server management in distributed data centers. In this process jobs are allocated according to the server choices. In this section server activation decides the active servers which are at slow time scale in the other section service rates of the section are controlled by the power scaling decisions. To solve this problem, we propose a novel approach based on decreasing time algorithm based on the priority values.

Keywords: Data Center, Power Management, Stochastic Optimization, Performance Analysis

I. INTRODUCTION

Over a few decades technology that insist computing in various aspects and improvement of those data center still in requirement phase. This can be stabilized by enabling a huge scale geographically distributed data centers with large numbers of servers. Data centers are centralized servers which organize storage and management in an effective manner. In which data centers compose lot of power for two usages such as running the actual equipment and cooling power of this equipments. In the current technical world every organization is using the data center which can be considered as brain of an organization. The serious discussion should be taken for improving the performance in hardware design and engineering as well as adjusting the CPU speed in a single server.

The Fig.1 shows the design of a data center; the consumption cost can be optimized by dealing the workloads in an effective manner. There are various algorithm and techniques have been in process dealing in controlling the workloads in several aspects. In this manner the known public data center is The National Climatic Data Center (NCDC) that holds the huge archive of weather information’s. But this section still in research with various issues as the computing demand has grown successfully. According to Carrie Higbie, of Global Network

Applications data center in every organization is under restriction access and the networking systems were controlled by automatic systems that monitors regularly overall server activity along with network performance and web traffic.

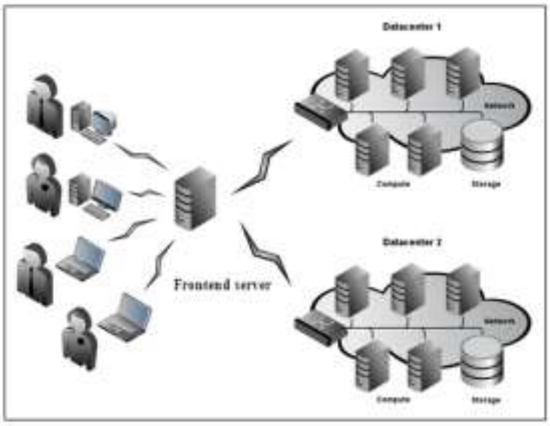


Fig.1 Data centers

While discussing this green computing was also popular among the organizations in this the computer user holds the overall responsibility. The aim of this approach is minimizing the energy and harmful materials. Cloud computing is a pattern that characterized the delivery and consumption of IT. Cloud computing consumes resource utilization which minimize the hardware equipments. It was a large set of server which is grouped with each other by means of internet. Here job scheduling plays vital role in arranging the job execution in an effective manner. The job allocation has to be faced various issues like which job is to be executed first, what resources to be allocated to do in a competent manner. The various servers may cause of network traffic which overall result on poor performances. A proper mechanism is to be prepared to over the issue that takes the performance of cloud computing into a new era.

II. RELATED WORKS

The importance of cloud centers and requirement of performance improvement grabs the attentions of various researchers in analyzing a prominent solution on these factors. From the beginning several peoples were developed some mechanism with the goal of improving performance, according to a report [1], Google (1 million), Microsoft (200K), Akamai (60-70K), INTEL100K), which has large number of various servers that are location in a various locations. In these the consuming cost are million dollars per year. With the goal of reduced power cost in data centers various works are carried out by [2][3][4][5][6][7] which can be discussed on two sections such as dealing with the saving power cost of DC power supplies, cooling systems and energy efficient chips. In the next section is sizing the data center as much as feasible.

The workload are major drawbacks to be controlled in data centers, to do this various algorithm are developed according to Lyapunov optimization concepts from [8] it designed with stochastic settings but this not suitable for real-time situations. Next on this way MapReduce-based application which makes use of web services that is indicated in [9]. By Gandhi et al. way developed certain management policies for minimizing the total power cap [11]. Some research on service providers like Google replicates data across were carried out on [12] for providing better services on I/O intensive jobs. According to Urgaonkar et al. he has discussed about the network utility optimization problems for job admission control, routing, and resource allocation [13].

Some algorithm were developed which are designed to work on single server with minimum power consumption subject to job deadlines and response time [14][15]. This work is continued on [16] in reducing the single server power consumptions. By the recent researches some works are implemented in making a better usage of power on data centers. According to El-Sayed et al. [17] and Liu et al. [18] introduce various strategies on cooling systems and they process renewable energy by making the server management together with cooling systems. In [19][20] they discussed about the existing techniques about the for reducing durable availability of MapReduce jobs on both the prevalence and the magnitude of task

III. EXISTING SYSTEM

The power cost minimization is the growing problem which is solved by using the SAVE (StochAstic power redUction schEme).This algorithm is categorized of three factors such as front-end rounting, backend server management and Queue update. It minimize the power consumption of the server and supports grren computing platform. It works on the basis of two scale mechanism that reduce the power cost in geographically distributed data centers. It serves job according to a particular slot even in queue due to high price at data center.

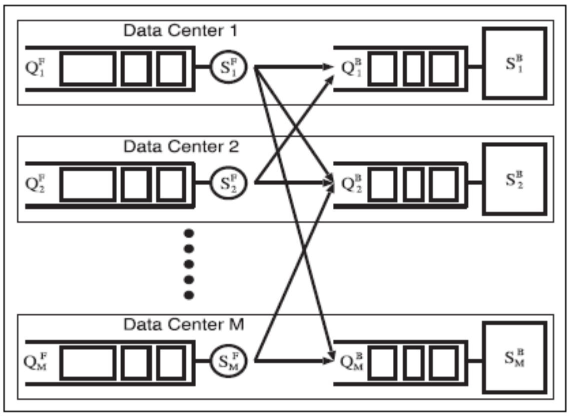


Fig.2 SAVE Mechanism

The mechanism sends the workload from front end to back end server this mechanism does not changed for that slot. For a larger time slot the algorithm optimized the time consumption in order to reduce the cost. This algorithm is based on Lyapunov optimization framework with queue stability. One of the best thing in this algorithm is inaccurate in queuing the backlog information which routes multiple jobs with lower power price. This mechanism is discreteness due to job sizes in which all tasks are assigned at back end cluster which is difficult to control the power consumption. The workload is cannot be solved by SAVE in an effective manner because some slots are failed to achive the power cost. Since in this approach the load balancing in all servers are not active because it assumes that servers can switch between active state and sleep state with same frequencies. The maximum workload is a greedy fashion that is not very effective in order to reducing power cost. The major drawback of this method is it require simultaneously activation and deactivation on multiple servers but it results in unsteadiness in power grids.

IV. PROPOSED SYSTEM

To overcome the problem as discussed in the existing system we proposed a decreasing time algorithm, initially it creates the priority list by means of it. In which it set the maximum priority to the very long task in order which completes the task in a shortest time period.

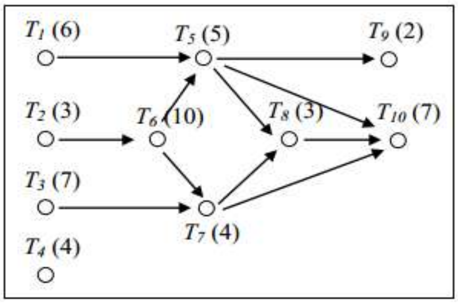


Fig.3 Proposed system

As shown in the Fig.3 the proposed system makes the priority list by arranging the tasks from longest time to shortest time in a prominent manner which will be more effective power consumption as better than SAVE method. With the Fig.3 the priority list is T6 (10), T3 (7), T10 (7), T1 (6), T5 (5), T4 (4), T7 (4), T2 (3), T8 (3), T9 (2). These compute a minimum time to complete the job along with highest total completion time. The difficulty is to identify which one is longest task that is high priority it could be practiced by preparing the list which is scheduled to process with the completion time of 32. Consider that this algorithm is processed with two processors finishes at time 32 with only 4 time units on the second processor. Such as from the list T3, T2, T4, T1, T6, T5

Machine 1: T3, T4, T5

Machine 2: T2, T1, T6

The remaining tasks are carried out in the same manner. Check that the sum of two sets at a value of 30. It can be mathematically explained by

$$((4/3)-1/(3m))T$$

In which T is the optimal time for schedule task and number of machine (two), is represented by m then the tasks are carried. The care should be taken in finding the critical path because it makes the process some time critical. The methodology behind this algorithm is the longer task sound good which complete the task quickly at the end the overall performance which results in reliable power consumption.

V. SYSTEM DESIGN

In this section the information are collected and servers are collected which are going to implement this process. A analysis should be taken what are the jobs and how it could be carried out. How the process is going too happened in geographically distributed datacenters.

Fig.4 Proposed system architecture

Based on this system the jobs are carried, in which the VM collects the overall task for a job of single server. In this the slot is prepared by implementing the decreasing time algorithm. Based on this a priority list is prepared by means of longest task in forward which was listed in times by decreasing order. Here two data centers are shown each one has clusters of systems which are connected to a server such as front-end proxy server and a back-end server cluster respectively. The workload is allocated by the user, so the workload arrived time are calculated and size of the workload are also noticed based on this analysis the above mentioned priority list is prepared for a slot. This is denoted by denoted by D = {D1,…, DM}, where the system operates in slotted time, i.e., t = 0, 1, 2, . . . with workload arriving time at Di by A(t) = (A1(t),…,AM(t)) denotes the arrival vector.

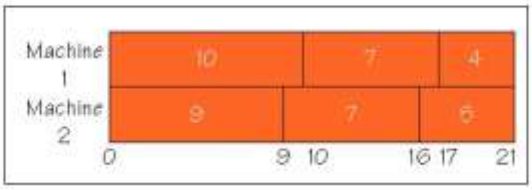


Fig.5 Task allocation

The above figure shows the how the priority list was carried out by two machines which results in optimum solutions.

VI. ALGORITHM IMPLEMENTATION

Step 1: Jobs allocating

Step 2: First identify the execution time of jobs before scheduling

Step 3: Calculate the priority values

Step 4: Fix the priority values.

Step 5: Based on the priority, we schedule the jobs

Step 6: Arbitrary List: A (6), B (5), C (7), D (2), and E (5)

Decreasing Time List: C (7), A (6), B (5), E (5), D (2)

Step 7: Execution of jobs

Step 8: Output the execution time

VII. RESULT & DISCUSSION

The jobs are arrived at the datacenters in which it is identified and sends to the router. The router schedules the job in which the backend cluster manipulates arrived jobs. The result shows list of arrived jobs and tables are scheduled. In which the task are analyzed among that the longest one in the topmost priority which can be seen in the Fig.6.

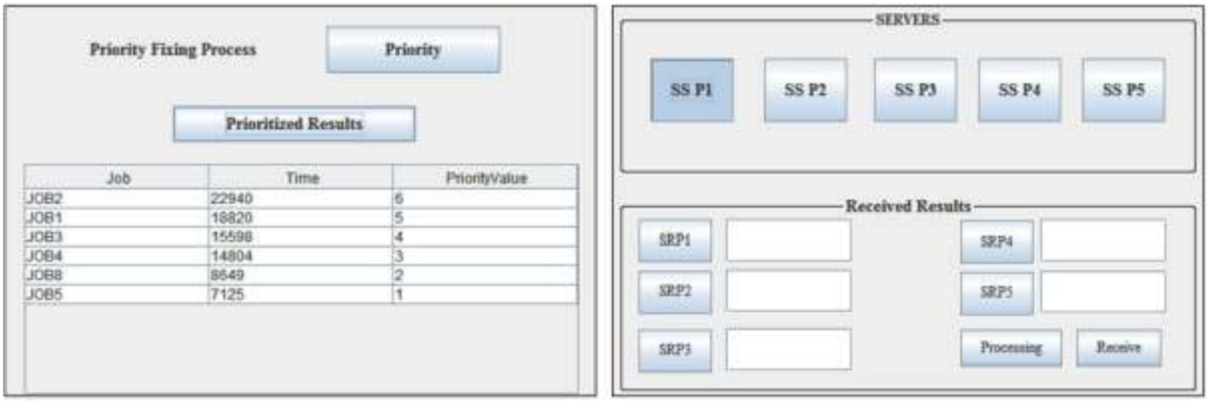


Fig.6 Server executes the jobs based on priority

After that the available server to which the execution jobs to be carried out. On that various job are allocated as per in the priority list with the help of host IP in each of the systems. Fig.7 shows some sample jobs such as converting image color and document in to PDF in a similar manner all these executions are traced and monitored prominently.

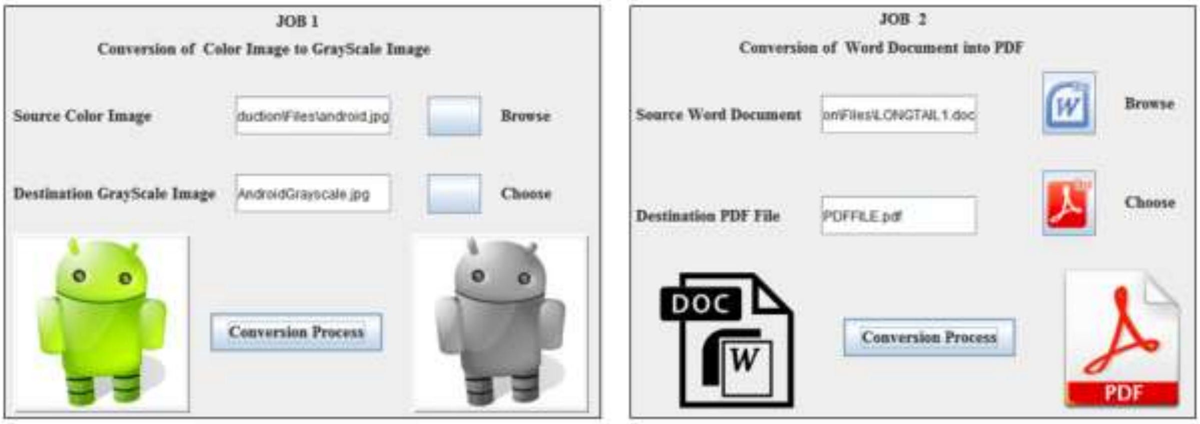
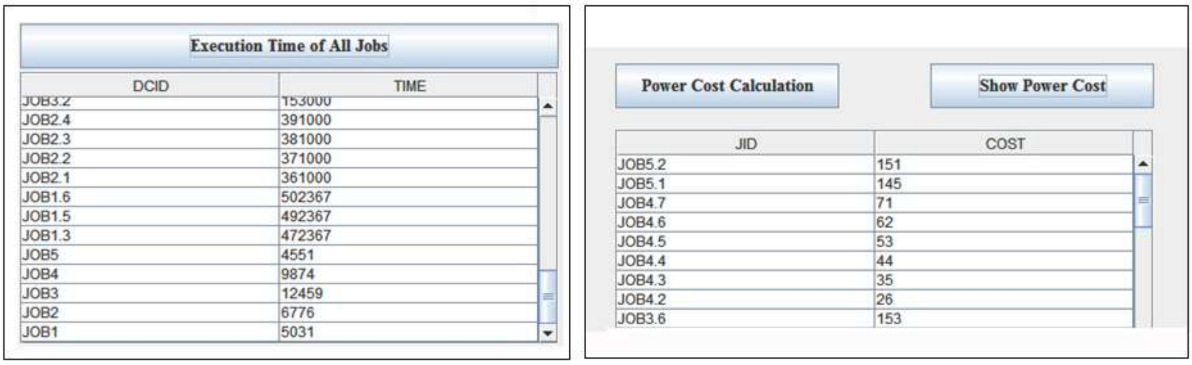


Fig.7 Conversion of color image to grayscale image and word document to pdf

By the Queue the jobs are executed and the results are tabulated as shown in the Fig.8 by the activation of server. This value is calculated in next stage to analyze the overall performance. In the same aspect the power consumption were also calculated which are also tabulated to analyze the final result.



Execution Time of All Jobs

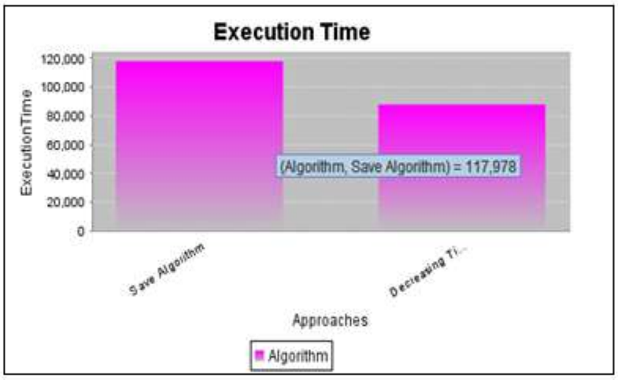
DCID	TIME
JOB3.2	153000
JOB2.4	391000
JOB2.3	381000
JOB2.2	371000
JOB2.1	361000
JOB1.6	502367
JOB1.5	492367
JOB1.3	472367
JOB5	4551
JOB4	9874
JOB3	12459
JOB2	6776
JOB1	5031

Power Cost Calculation   Show Power Cost

JID	COST
JOB5.2	151
JOB5.1	145
JOB4.7	71
JOB4.6	62
JOB4.5	53
JOB4.4	44
JOB4.3	35
JOB4.2	26
JOB3.6	153

Fig.8 Execution time and power cost calculation

Based on these two reading a graphical representation is created. The Fig.9 justified the decrease the time algorithm is more feasible when compared to result attained by the SAVE algorithm. It is clear how the tasks of the jobs are carried out and power cost taken by each jobs are graphically shown, as prominent support for proving the efficiency of our proposed algorithm. Thus proves that minimum time consumption is result in minimum power cost consumption which was successfully achieved by our proposed algorithm.



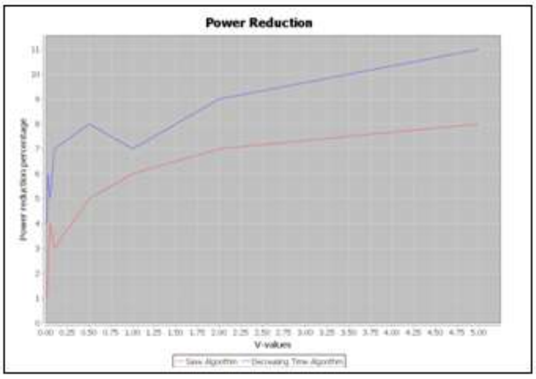


Fig.9 Graphical representation of execution time and power reduction between SAVE algorithm and decreasing time algorithm

VIII. CONCLUSION & FUTURE WORK

In this paper we shown how our proposed system is carried out the performance done is more efficient than earlier systems. Here it clearly explained that decreasing time algorithm based priority list is how effective than the SAVE system in achieving the minimum power cost consumptions. In addition to it we also show that how our approach is effectively handling the delay tolerant workloads as well as the network traffic in distributed data centers. Finally we provide a prominent solution for doing activation and deactivation of multiple servers to do process. The overall result based on real-time data in simplifying the problem by assuming the processing time of each job is proportional to the amount of work. In future the work is carried out in improving the performance of datacenters by achieving more accuracy with the help of modern algorithm that satisfies the delay tolerance.

REFERENCES

[1] www.gizmodo.com/5517041/, 2013.

[2] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, “Cutting the Electric Bill for Internet-Scale Systems,” Proc. ACM SIGCOMM Conf. Data Communication (SIGCOMM ’09), 2009.

[3] Z. Liu, A. Wierman, S. Low, and L. Andrew, “Greening Geographical Load Balancing,” Proc. ACM SIGMETRICS Joint Int’l Conf. Measurement and Modeling of Computer Systems (SIGMETRICS ’11), 2011.

[4] M. Lin, A. Wierman, L. Andrew, and E. Thereska, “Dynamic Right-Sizing for Power-Proportional Data Centers,” Proc. IEEE INFOCOM, 2011.

[5] A. Wierman, L. Andrew, and A. Tang, “Power-Aware Speed Scaling in Processor Sharing Systems,” Proc. IEEE INFOCOM, 2009.

[6] R. Stanojevic and R. Shorten, “Distributed Dynamic Speed Scaling,” Proc. IEEE INFOCOM, 2010.

[7] A. Gandhi, V. Gupta, M. Harchol-Balter, and A. Kozuch, "Optimality Analysis of Energy-Performance Trade-Off for Server Farm Management," Performance Evaluation, vol. 67, pp. 1155-1171, Nov. 2010.

[8] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queueing Systems & Scheduling Policies for Maximum Throughput in Multihop Radio Networks," IEEE Trans. Automatic Control, vol. 37, no. 12, Dec. 1992, pp. 1936-1949.

[9] A.K. Mishra, J.L. Hellerstein, W. Cirne, and C.R. Das, "Towards Characterizing Cloud Backend Workloads: Insights from Google Compute Clusters," SIGMETRICS Performance Evaluation Rev., vol. 37, no. 4, Mar. 2010, pp. 34-41.

[10] Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, and N. Gautam, "Managing Server Energy and Operational Costs in Hosting Centers," Proc. ACM SIGMETRICS Int'l Conf. Measurement and Modeling of Computer Systems (SIGMETRICS '05), 2005.

[11] A. Gandhi, M. Harchol-Balter, R. Das, and C. Lefurgy, "Optimal Power Allocation in Server Farms," Proc. 11th Int'l Joint Conf. Measurement and Modeling of Computer Systems (SIGMETRICS '09), 2009.

[12] http://googleenterprise.blogspot.com/2010/03/disasterrecovery- by-google.html, 2013

[13] R. Urgaonkar, U. Kozat, K. Igarashi, and M. Neely, "Dynamic Resource Allocation and Power Management in Virtualized Data Centers," Proc. IEEE Network Operations and Management Symp. (NOMS), 2010

[14] F. Yao, A. Demers, and S. Shenker, "A Scheduling Model for Reduced CPU Energy," Proc. 36th Ann. Symp. Foundations of Computer Science, 1995.

[15] K. Pruhs, P. Uthaisombut, and G. Woeginger, "Getting the Best Response for your Erg," ACM Trans. Algorithms, vol. 4, July 2008, pp. 1-17.

[16] S. Albers, "Energy-Efficient Algorithms," Comm. ACM, vol. 53, May 2010, pp. 86-96.

[17] N. El-Sayed, I.A. Stefanovici, G. Amvrosiadis, A.A. Hwang, and B. Schroeder, "Temperature Management in Data Centers: Why Some (Might) Like It Hot," Proc. ACM SIGMETRICS, 2012.

[18] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser, "Renewable and Cooling Aware Workload Management for Sustainable Data Centers," Proc. 12th ACM SIGMETRICS/PERFORMANCE Joint Int'l Conf. Measurement and Modeling of Computer Systems (SIGMETRICS '12), 2012.

[19] M. Zaharia, D. Borthakur, J. Sen Sarma, K. Elmeleegy, S. Shenker, and I. Stoica, "Delay Scheduling: A Simple Technique for Achieving Locality and Fairness in Cluster Scheduling," Proc. Fifth European Conf. Computer systems (EuroSys '10), 2010.

[20] G. Ananthanarayanan, S. Kandula, A. Greenberg, I. Stoica, Y. Lu, B. Saha, and E. Harris, "Reining in the Outliers in Map-Reduce Clusters Using Mantri," Proc. Ninth USENIX Conf. Operating Systems Design and Implementation (OSDI '10), 2010.