ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, Vol 8(35), DOI: 10.17485/ijst/2015/v8i35/84044, December 2015 ISSN (Online) : 0974-5645

An Energy Efficient Data Centre Selection Framework for Virtualized Cloud Computing Environment

B. Santhosh Kumar[1]* and Latha Parthiban[2]

[1]Department of CSE, G. Pulla Reddy Engineering College, Kurnool – 518007, Andhra Pradesh, India; santhoshkumar.bala@gmail.com
[2]Department of CSE, Community College, Pondicherry University, Pondicherry – 605014, Tamil Nadu, India; lathaparthiban@yahoo.com

Abstract

Background: Data centers are the major sources for implementing critical issues in a cloud computing environment like networking, task scheduling, storage and the distribution of data. The objective is to design a data centre selection framework for submission of tasks to a data centre with minimal energy consumption. **Methods:** A framework is proposed to select a data centre for tasks submission with minimum total energy consumed by the servers, computer room air conditioning units and other IT equipments like routers, switches. The service provider has to register all the data centers in a data centre registry and the data centre selection framework decides upon task allocation to the appropriate data centre with minimal energy consumption. The results are simulated using Google Cloud Trace logs and cloud analyst tool. The proposed work is compared with shortest distance first and round robin methods. **Findings:** The performance analysis indicate that the proposed framework is an efficient way for tasks submission to a data centre as it concentrates on minimizing the energy consumption which in turn reduces operational expenditures of the cloud environment. The comparisons of the work with other algorithms indicate that the proposed framework consumes far less energy. **Improvements:** The framework can be implemented by tools like IBM Bluemix which uses Internet of Things (IoT) technology for reading the energy using sensors.

Keywords: Cloud Computing Environment, Data Centre Selection Framework, Energy Consumption, Tasks Submission

1. Introduction

Cloud computing is the on growing prominent technology which offers services to customers on a “pay-as-you-go” model[1]. The Gartner Report forecasts that the cloud services are expected to grow into a $210 billion market by the year 2016[2]. Many of the IT companies can concentrate on innovation and development while the service providers care about software and hardware infrastructures[3]. Data centres play a vital role in delivering the services in the cloud. Many of the critical issues like task processing, storage and distribution of data are taken care by the data centres. The techniques adopted to choose a data centre for tasks submission plays a crucial role in cloud environment. Some of the predominant ones are mentioned below:

1.1 Spot Electricity Prices

The electricity consumed by the data centres are taken into account for choosing a data centre. The future variations in the prices are predicted and the data centre which consumes minimum electricity is given preference over others. The main disadvantage of this procedure is that it takes into account a lot of energy consumed by the elements which are in idle state and also some of the non crucial elements like tube lights etc., become a part of price prediction.

*Author for correspondence

1.2 Shortest Distance First

This approach calculates the distance from where the requests have been originated and forwards them to the closest distant data centre. The drawback of this approach is that it doesn't take into consideration the energy factor which is quite expensive when compared to the network elements that have to be established to route the requests to data centres. Placing a few more elements like routers, bridges etc., will expedite the process of request forwarding even if a data centre is located at a far place.

1.3 Round Robin Technique

This procedure will randomly choose a region for servicing the tasks and a data centre in that region is selected for processing. This approach might lead to a delay factor and it doesn't consider energy factor which results in large operational expenses.

To overcome the shortcomings in the above procedures and to consider the energy consumption of the operational and network elements of a data centre we try to propose a framework which calculates the total energy provided by all the data centres so that the incoming requests can be allocated to data centres with minimum energy consumption. The idle status of a server is brought down by employing virtualization effectively and the energy consumed by the virtual machines is only considered for calculation of energy consumption. The energy of networking elements like switches, routers and electrical equipments like air conditioning units, chillers are taken in to consideration.

2. Contribution

The major contributions of the paper are as follows:

- A Data Centre Selection Framework (DCSF) is proposed which submits batch of tasks to a data centre with minimum energy consumption.
- The concept of Data Centre Registry (DCR) is proposed where the service provider can register the data centre information

A lot of research work has been done to select a data centre for tasks submission. Shuyi Yan et al.[4] proposed data centre selection based on two policies steady state probability and network state policy where probability is calculated based on location and network state. A matrix based data centre selection approach has been formulated by Amol Jaikar et al.[5] by considering the distance and cost factors in a matrix format. Deepak Kapgate et al.[6] developed a service broker algorithm where proximity of data centre is considered as a prime factor for data centre selection. The cost based data centre selection for large scale networks has also been proposed in the literature[7].

Myint and Thandar Thein[8] proposed a machine learning based electricity price prediction for data centre selection. Hong Xu and Baochun Li[9,10] formulated a sub gradient method and Nash bargaining procedure for selection of data centre. Round robin algorithm for service proximity has been proposed by Vaishali Sharma et al.[11] to select a data centre in a single region. Devyaniba Chudasama et al. proposed a cost effective selection of data centre by proximity based routing policy[12]. Unfortunately, to the best of our knowledge all the proposed works concentrate on selection of data centre based on prediction criteria, spot electricity prices and proximity based approach. The current energy consumption of a data centre has not been taken in to account for data centre selection. In this proposed work, we focus on the consumption of energy of the active elements of a data centre and try to develop a framework which calculates the total energy consumed and forward the tasks generated to the appropriate data centre.

3. Proposed Work

Generally the tasks submitted by the users are routed through web servers to the data centers of service providers. In the proposed work a Data Centre Selection Framework (DCSF) is designed so as to process the tasks. A User Base (UB) is considered to be a region with many users trying to submit different tasks for processing at data centers. The tasks are forwarded from the web servers to the DCSF which in turn takes the decision of submitting the tasks to appropriate data centre with minimal energy consumption. The entire process can be depicted using Figure 1.

A unique Id will be generated for all the data centres associated with a service provider. All the data centres have to be registered with a centralized repository referred as Data Centre Registry (DCR). All the information related to the data centres like Data Centre Id, No. of Servers, Switches, Routers, Computer Room Air Conditioning Units (CRACs), Chillers and the average energy consumed by each component over the past few days can

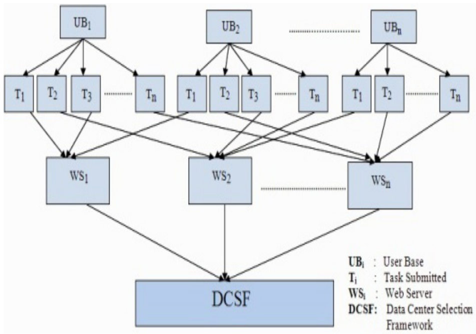

Figure 1. Submission of tasks to DCSF.

be updated. A Data Centre Id (DCID) is a 16 bit number in which the first 4 bits are associated with service provider Id and the remaining 12 bits are allotted to data centre id which should be a unique number generated by the service provider. An overview of DCR fields can be seen in Table 1.

Table 1. Contents of data centre registry

Data Center Id (DCID)	No. of Servers (NS)	No. of Switches (NSW)	Energy Consumed Per Switch (E_{SW})	No. of Routers (NRT)	Energy Consumed Per Router (E_{RT})	No. of CRACs (NCRAC)	Energy Consumed Per CRAC (E_{CRAC})	No. of Chillers (NCH)	Energy Per Chiller (E_{CH})

3.1 Data Centre Selection Framework

The Data Centre Selection Framework (DCSF) is illustrated in Figure 2. It consists of different modules whose output put together will determine the data centre for which all the tasks should be submitted. All the tasks which are submitted to DCSF are stored in a Waiting Queue. If the Ready Queue is empty then all the tasks are transferred from Waiting Queue into Ready Queue. If the no. of tasks (λ) exceeds some Threshold value (Th) or the Time > 60s whichever is earlier then the tasks are assigned to Task Allocator (TA). Then TA sends a request to Data Centre Selector (DCS) to decide upon the data centre for which the tasks should be submitted. DCS can be organized in to three different modules:

3.1.1 Server Energy Consumption Module (SECM)

The virtualization technology is adopted to make the processing of data centres faster and efficient. Many works have been proposed on using virtualization effectively in data centres[13-18]. But the proposed work is based on the energy consumed by the servers that can be calculated by the total energy consumed by executing the tasks on virtual machines[19].

Let S = {S_1, S_2, S_3… S_n} be the set of servers located in a data centre DC_i. For each server S_k Let VM_k = {VM_{1k}, $VM_{2k...}$, VMnk} be the set of virtual machines created. A decision variable x_{ijk} is set to 1 if task T_i is allocated to virtual machine VM_{jk}. Let EC_{ijk} be the energy consumed by task T_i running on VM VM_{jk} and let ET_{ijk} be the execution time. The energy consumption rate of the VM is denoted by ECR_{ijk} and energy consumption EC_{ijk} can be calculated as follows:

$$EC_{ijk}=ECR_{ijk}\times ET_{ijk} \quad (1)$$

Where $ET_{ijk} = ft_{ijk} - st_{ijk}$ (finish time – start time)
Hence using (1) the total energy consumption can be calculated as follows:

|S| |VMk| |T|
TE1 = Σ Σ Σ xijk × ECijk k=1 j=1 i=1
|S| |VMk| |T|
TE1 = Σ Σ Σ xijk × ECRijk × ETijk (2)
k=1 j=1 i=1

Where |T| Indicates Total no. of tasks submitted to VM

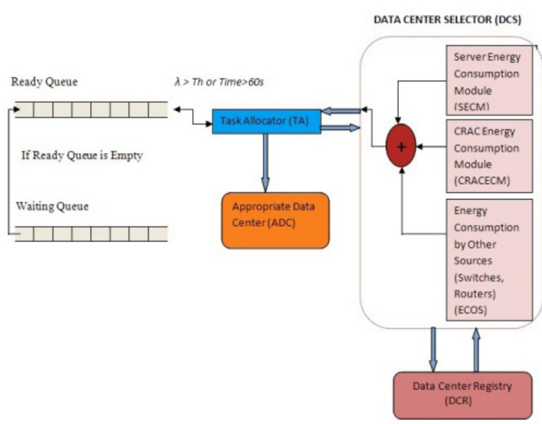

Figure 2. DCSF for allocation of tasks to selected data centre.

3.1.2 CRAC Energy Consumption Module (CRACECM)

The No. of AC devices, chillers and the energy consumed per device can be retrieved from data centre registry and the total energy consumed can be calculated as follows:

$$TE_2 = NCRAC \times E_{CRAC} + NCH \times E_{CH} \quad (3)$$

3.1.3 Energy Consumption by Other Sources (ECOS)

The No. of Switches, routers and their energy consumption per device can be retrieved from data centre registry and the total energy consumed can be calculated as follows:

$$TE_3 = NSW \times E_{SW} + NRT \times E_{RT} \quad (4)$$

The output of the three modules will be the total energy consumed by the data centre which can be given by the following equation:

$$E_{TOTAL} = TE_1 + TE_2 + TE_3. \quad (5)$$

The total energy consumed by all the data centers is calculated by the DCS and the task allocator selects DCID with the minimum total energy consumption i.e. $E_{TOTAL.}$ Then all the tasks are submitted to the Appropriate Data Centre (ADC) with minimum E_{TOTAL} for further processing. The algorithm stating the entire process is given below:

3.2 Data Centre Selection Algorithm START

For Each Service Provider SP
Register Each Data Centre DC_i in Data Centre Registry DCR as in Table 1
End For
Initialize Ready Queue RQ←Null,
Size (RQ)← Threshold Value (Th),
Waiting Queue WQ←Null
For Each Task T_i submitted by User Base UB_i
Assign WQ←T_i
If RQ is Null Then Assign RQ←WQ
Until no. of tasks (λ) > Size (RQ)
End If
End For
If λ > Size (RQ) or Time > 60s
Request Task Allocator TA to find Appropriate Data Centre ADC.TA assigns the request to DCS which retrieves each T_i from RQ
End If
DCS finds ADC as follows:
|S| |VMk| |T|
TE1 = Σ Σ Σ xijk × ECRijk × ETijk
K = 1 j = 1 i = 1
TE2 = ECRAC+NCH * ECH
TE3 = ESW+ NRT*ERT
$E_{TOTAL} = TE_1 + TE_2 + TE_3$
ADC←DCID (min (E_{TOTAL}))
For Each Task $T_{i,}$ TA do
Assign ADC ← T_i for Processing
End For
End

4. Performance Analysis

Simulation has been performed by considering the cloud analyst tool which is based on cloud sim. It allows defining the data centre elements and choosing data centres across different parts of globe like Asia, North America etc.. To evaluate the proposed algorithm the real world trace referred in Google Cloud Trace logs has been considered[20]. It contains information of 25 million tasks that emerged for 29 days. As it is impossible to conduct the analysis on all the tasks only the tasks submitted to the cloud system in the first 5 hours of the day 18 are taken in to consideration. The count of the tasks submitted in every 60 seconds in those 5 hours is depicted using Figure 3. During this time period the count of the tasks is over 200 thousand.

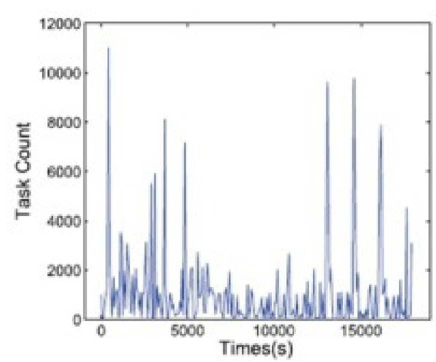

Figure 3. Submission of tasks to cloud.

The start and finish time of the tasks are recorded in the trace logs from which we obtained the execution time. The sample data which is further required for our simulation is assumed as given in Table 2. The Data Centre Characteristics class of cloud analyst is coded with all the variables as defined in the table. For our convenience data

Table 2. Data assumption for simulation

Data Center Id (DCID)	No. of Servers (NS)	No. of Switches (NSW)	Energy Consumed Per Switch (E_{SW})	No. of Routers (NRT)	Energy Consumed Per Router (E_{RT})	No. of CRACs (NCRAC)	Energy Consumed Per CRAC (E_{CRAC})	No. of Chillers (NCH)	Energy Per Chiller (E_{CH})
DC1	45	300	2.5	35	1.5	24	4.5	20	3.2
DC2	32	225	2.2	23	1.75	22	4.2	22	2.7
DC3	48	143	2.7	28	1.9	25	4.1	25	3.5
DC4	38	345	2.1	32	2.1	32	3.8	26	2.9
DC5	40	173	2.9	27	2.3	35	4.2	29	2.8

centre Ids are not assumed in binary format but are represented as DC1... DC5.

The energy consumption is represented in KWh. The data centers are assumed to be in different parts of the world which will vary in the no. of cooling units and chillers used depending on the weather conditions. The proposed algorithm is compared with the other existing algorithms like Shortest Distance, Round Robin and results are analyzed. The metrics taken for analysis of the algorithm are Total Energy Consumption (TOT) and Energy Consumption Per Task (ECT).

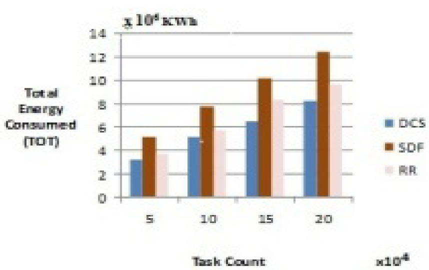

Figure 4. Performance analysis using TOT metric.

The proposed algorithm Data Centre Selection (DCS) is coded as part of the Data Centre Class in cloud analyst and compared with the Shortest Distance First (SDF) and Round Robin (RR) algorithms. The results are depicted in Figure 4 and Figure 5. The results indicate that the proposed work requires far less energy than the other existing approaches to complete the tasks. To implement SDF the data center DC4 is taken as the shortest distance from the place where the tasks originated.

5. Conclusion and Future Work

In this paper we proposed a framework for selection of data center based on energy efficiency. This approach mainly concentrates on the submission of tasks to energy efficient data center which in turn results in minimizing the operational expenditures of the cloud environment.

The work can be further extended by simulation of results using tools like IBM Blue Mix where the energy of the data centers can be simulated by IOT sensors. Efforts are being made to implement a cloud environment in our premises which will incorporate this framework in the backend processing so that it can be easily adapted by any other applications that are developed in cloud environment.

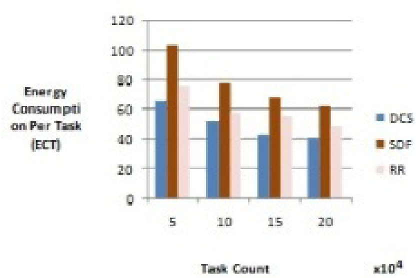

Figure 5. Performance analysis using ECT metric.

6. References

1. Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I. Cloud computing and emerging IT platforms: Vision, hype and reality for delivering computing as the fifth utility. Future Generation Computer Systems. 2009; 25(6):599–16.
2. Gartner. Forecast overview: Public cloud services. Worldwide, 2013-2019, 1Q15 Update. 2013. p. 58–71.
3. Greenpeace international. Make IT green, cloud computing and its contribution to climate change. 2010. p. 128–31.
4. Yan S, Wang X, Razo M , Tacca M, Fumagalli A. Data centre selection: A probability based approach. Proceedings of 16th International Conference on Transparent Optical Networks (ICTON); Graz. 2014. p. 1–5.

5. Jaikar A, Kim G-R, Noh S-Y. Matrix-based data centre selection algorithm for a federated cloud. International Journal of Multimedia and Ubiquitous Engineering. 2014; 9(6):153–8.
6. Kapgate D. Efficient service broker algorithm for data centre selection in cloud computing. International Journal of Computer Science and Mobile Computing. 2014; 3(1):355–65.
7. Rekha PM, Dakshayini M. Cost based data centre selection policy for large scale networks. Proceedings of 2014 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC); Chennai. 2014. p. 18–23.
8. Swe K, Myint S, Thein T. Energy and cost efficient data centre selection for cloud services. Proceedings of 3rd International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2014); Singapore: PSRC. 2014. p. 35–8.
9. Xu H, Li B. A general and practical data centre selection framework for cloud services. Proceedings of 2012 IEEE 5th International Conference on Cloud Computing (CLOUD); Honolulu, HI. 2012. p. 9–16.
10. Xu H, Li B. Cost efficient data centre selection for cloud services. Proceedings of 2012 1st IEEE International Conference on Communications in China (ICCC); Beijing. 2012. p. 51–6.
11. Sharma V, Rathi R, Bola SK. Round-robin data centre selection in single region for service proximity service broker in cloud analyst. International Journal of Computers and Technology. 2013; 4(2):254–60.
12. Chudasama D, Trivedi N, Sinha R. Cost effective selection of data centre by proximity-based routing policy for service brokering in cloud environment. International Journal of Computer Technology and Applications. 2012; 3(6):2057–9.
13. Goiri I, Berral JL, Fito JO, Julia F, Nou R, Guitart J, Gavalda R, Torres J. Energy-efficient and multifaceted resource management for profit-driven virtualized data centres. Future Generation Computer Systems. 2012; 8(4):718–31.
14. Jiankang D, Hongbo W, Yangyang L, Shiduan C. Virtual machine scheduling for improving energy efficiency in Iaas cloud. China Communications. 2014; 11(3):1–12.
15. Shyamala K, Sunitha RT. An analysis on efficient resource allocation mechanisms in cloud computing. Indian Journal of Science and Technology. 2015; 8(9):814–21.
16. Ravi TP, Khadar BSK. An evolutionary computing based energy efficient VM consolidation scheme for optimal resource utilization and QoS assurance. Indian Journal of Science and Technology. 2015; 8(26):1–11.
17. Xiao Z, Song W, Chen Q. Dynamic resource allocation using virtual machines for cloud computing environment. IEEE Transactions on Parallel and Distributed Systems. 2013; 24(6):1107–17.
18. Papagianni C, Leivadeas A, Papavassiliou S, Maglaris V, Cervello-Pastor C, Monje A. On the optimal allocation of resources in cloud computing Networks. IEEE Transactions on Computers. 2013; 62(6):1060–71.
19. Zhu X, Yang LT, Chen H, Wang J, Yin S, Liu X. Real-time tasks oriented energy-aware scheduling in virtualized clouds. IEEE Transactions on Cloud Computing. 2014; 2(2):168–80.
20. Google Cluster Data V2 [Internet] 2011. Available from: http://code.google.com/p/googleclusterdata/Wiki/ClusterData2011_1