Journal of Computer Science

Original Research Paper

Multi Resources and Constraints Based Consolidation Model to Mitigate the Energy Extravagance

Archana Patil and Rekha Patil

Department of Computer Science and Engineering, PDA College of Engineering, Kalaburagi, India

Article history
Received: 12-04-2022
Revised: 27-06-2022
Accepted: 29-06-2022

Corresponding Author:
Archana Patil
Department of Computer Science and Engineering, PDA College of Engineering, Kalaburagi, India
Email: archanbpatil@gmail.com

Abstract: To design green clouds with optimal energy and resource consumption, the host consolidation process became the most prominent cloud management policy today. In case of a consolidation, the underloaded host's VMs are shifted to the other hosts to control the energy extravagance at a data center. The consolidation process is considered an NP-hard problem, as it depends on several computational resources (i.e., CPU, RAM, Bandwidth) and migration constraints. Although many former cloud researchers were focused on designing the consolidation process, the total consolidation process is based on single computational resource (i.e., CPU) utilization only. In a single resource-based consolidation model, only the selected resource utilization value is considered in decision-making and the other resource utilization values are simply discarded. The problems identified in single resource-based consolidation process are: (i) Ignoring the other prominent resources in consolidation (ii) unnecessary consolidation of hosts (iii) aggressive migrations and (iv) energy extravagance. To overcome the limitations involved with the single resource-based consolidation process, this study proposed the "Multi Resources and Constraints-based Consolidation Model (MRCCM)". The MRCCM is designed to solve the energy extravagance issues and aggressive migrations in host consolidation process. In MRCCM, multiple resources and resource constraints are considered while host consolidation for best target host selection and efficient resource management. Cloudsim toolkit with 800 HP-ProLiant host models and EC2 VM models are configured for experimental analysis. Proposed MRCCM model with MMT host selection method recorded the total 11570 migrations, 1853 shut_downs and 111.84 Kwh of energy consumption. Experimental results are proven that the MRCCM controlled the energy extravagance, unnecessary consolidations, and aggressive migrations at a considerable rate when compared to the other prominent consolidation models.

Keywords: MRCCM, VM Migration, Host Consolidation, Load Balancing, Green Cloud Computing

Introduction

For decades the growth rate of cloud adoption for enterprises was scaling up at a massive rate and is estimated to reach 623.3 billion dollars by 2023. As the world is moving toward green computing Gangadharan and Murugesan (2012), building green clouds with energy and resource-efficient solutions has become a major challenging task today. Efficient utilization of energy and resources will assure the reduction of carbon emissions (CO_2) to the environment. Among the all consumptions of the cloud, energy is too prominent and highly consuming resource. An inefficient model of energy consumption in the cloud will increase the power bills and causes to release the of greenhouse gases into the environment. Hence there is a need to concentrate on building energy-efficient clouds.

Recently the virtualization models by Uhlig *et al.* (2005) and Xing and Zhan (2012) were introduced with the cloud, to utilize the physical hardware resources efficiently by mapping them with logical virtual machines. In Sabahi's (2012) model, the hypervisor does act as a mediator between the hardware and virtual layer to create multiple virtual machines on a single physical host. Compared to the former cloud models, virtualization increases resource utilization and reduces energy consumption. Although virtualization made the cloud viable computing (Saranya and Fatima, 2021) paradigm in terms of cost, speed, and utilization, it is still suffering from the energy extravagance happening in the cloud, due to the imbalance of the load distribution among the VMs. VM migration and consolidation operations by Ahmad *et al.* (2015) and Li *et al.* (2016) are proven as the best solutions for load balancing at cloud data centers. Migration will shift the VMs from the overloaded host to others, to avoid the latency in processing the user jobs and to balance the load among the hosts. By balancing the load, migration helps in increasing the speed and reliability of a data center. Similarly, the consolidation process will shift the VMs from the underloaded hosts to the others and pause or shut downs the underloaded hosts avoiding the energy extravagance of Beloglazov and Buyya (2010). The consolidation process controls the energy extravagance by pausing or shutting downing the under-subscribed hosts and assures the high-end utilization of resources, which are allocated to the host.

Although the migration and consolidation techniques have become reliable solutions for load balancing are suffering from some considerable limitations: Dynamic load balancing, optimal threshold selection, energy efficiency (Patel and Makwana, 2016), and multi-resource-aware consolidation. Among these limitations, an optimal threshold selection plays a vital role in both migration and consolidation processes. In 2020, as part of their research on "green cloud computing" Patil and Patil (2020) proposed the optimal threshold selection algorithms for migration and consolidation processes. Their proposed SLAPT and DIQR models not only addressed the optimal threshold selection issue but also justified the energy efficiency, dynamic load balancing, and SLA violation too.

According to Minas and Ellison (2009) and Park and Pai (2006), the real-time energy consuming statistics at data centers, the under-subscribed (loaded) hosts of the data center are consuming the energy, which is nearly equal to the over or adequately loaded hosts. It means the under-subscribed hosts cause energy extravagance at a high level because these are consuming equal energy by processing very less or no jobs. At regular intervals, these underloaded hosts must be identified and consolidated to control the energy extravagance at data centers. In the consolidation process, the under-subscribed hosts are initially identified and tagged and later their VMs are shifted to the other hosts to free them from energy consumption.

Generally, the physical hosts are equipped with several resources (i.e., CPU, RAM, BW, and Storage) in the cloud, which are later distributed among the logical virtual machines for job processing. From the former research works on load balancing with migration and consolidation, it is identified that most of the works proposed the migration and consolidation techniques based on a single resource (i.e., CPU) utilization value. In reality, the other resources (i.e., RAM, BW, and Storage) also have considerable importance in job processing along with CPU. Beloglazov and Buyya (2010) proposed the energy-efficient consolidation process based on optimal threshold values. The model discussed several VM selection and allocation models along with the threshold selection strategies. In the end, scholars discussed the importance of implementing the multi-resource (i.e., RAM, BW, and Storage) based consolidation at data centers. Zhang *et al*. (2012) proposed the constraints-based migration model for virtual machines to avoid aggressive migrations in the cloud. In that model, the scattered migration algorithm is reduced to control the number of migrations while balancing the load. Finally, their plan is designed to conduct migration experiments, using multiple resources to control aggressive migrations. Yu *et al*. (2020) proposed the dynamic load balancing scheme, which implements the migrations based on stochastic characterizations. To overcome the migration overhead, scholars implemented the multi-resource-based migration, which evaluates the network Band Width (BW) along with the CPU time. Khan *et al*. (2018) conducted a review on consolidation algorithms and outlined the research areas of consolidation. The authors specified that only the CPU-based consolidation methods will saturate the CPU utilization, but not the other resources like memory and bandwidth. Ferdaus *et al*. (2014) proposed the VM consolidation process using the ACO with meta-heuristic constraints to avoid the delay in decision makings of the consolidation process. The aforementioned research scholars on the cloud projected the multi-resource-based consolidation as their future work.

Single resource-based load balancing (migration or consolidation) methods cannot distribute the load evenly and fails in satisfy the other resource requirements, which leads to aggressive migrations. The major limitations involved with single resource-based consolidation are: (i) Prominent and effectible resources are discarded in the consolidation decision-making process (ii) Inappropriate selection of hosts for consolidation (iii) aggressive and erroneous migrations and (iv) energy extravagance. To overcome the limitations involved with single resource-based consolidation, this study proposed the "Multi Resources and Constraints-based Consolidation Model (MRCCM)", to implement an efficient consolidation process, which is energy and resource-aware. Resource-aware consolidation was proposed to reduce the aggressive migrations of VMs. With the help of the Cloudsim toolkit by Huang and Tsang (2014) and the planet lab dataset, the experiments were conducted on MRCCM and the results proved the efficiency of the MRCCM compared to other prominent consolidation models.

This study concentrated on designing the efficient consolidation model based on multiple resources and constraints. The paper is organized as follows: Related works described the former research works and the basics of consolidation. MRCCM section explained the proposed consolidation model and the experiments section outlined the comparative analysis of MRCCM.

Related Work

In cloud computing, a data center is a collection of physical hosts, which are equipped with several resources (i.e., CPU, RAM, BW, HD, etc.) to connect, process, and store the user application data. A virtual layer with VMs is implemented on physical hosts to create the logical servers and to utilize the resources efficiently. To process the workload (i.e., user jobs) on time and to avoid overloading issues, the incoming workload should be evenly distributed among the VMs in a decentralized manner. While the execution of the jobs, the adaptive dynamic upper threshold (δ_U) and the lower threshold (δ_L) values will be defined to balance the load at runtime using the migration and consolidation processes. The prominent resources to be considered for load balancing at the data center level are categorized and presented in Fig. 1. As discussed, the threshold value plays a vital role in dynamic migration and consolidation operations. After the threshold values calculation, the next level is designing a model for consolidation with several policies and constraints. An inefficient model of migration and consolidation leads to several issues that were discussed in the introduction section.

This section explored the main objectives to consider while designing the dynamic consolidation model at data centers. As shown in Fig. 1, the considerable objectives for consolidation are allocated resources, VM selection policies, target constraints, and relevant algorithms. The consumption value of the resources allocated to the host (i.e., CPU, RAM, BW, and HD) are mainly used to determine the underloading and overloading problems. Most of the former researches accounted only for the CPU consumption value but not the other resources, although they have equal importance in processing the user tasks. Hence it is recommended to consider the multiple resources in consolidation decisions. Once the underloaded hosts are selected for consolidation, the next step will be selecting the target hosts for shifting the VMs from the under-loaded host. At this time various selection policies are to be considered to assure the smooth execution of consolidation without any problems in further. The popular target host selection (VM selection) policies by Beloglazov and Buyya (2012) are migration time-based selection, correlation-based selection, dynamic inter quartile range-based selection, and random selection. Along with these VM selection policies, the target host selection constraints should be considered in the consolidation process to satisfy the end user opted agreements while servicing subscriptions. Resource awareness is the main constraint to follow while consolidation to satisfy end-user agreements. Energy and network awareness-related constraints will benefit the service provider with more profits, by saving energy and utilizing the resources efficiently. Finally, this whole consolidation process should be framed into the best suitable algorithms. Ferdaus *et al.* (2014) considered that the consolidation process is an NP-hard problem, to solve this problem the best class algorithms with optimal solutions are heuristic, meta-heuristic, and hybrid meta-heuristic algorithms. By insisting the all possible objectives in the consolidation model, it's possible to address the frequently occurring limitations of the consolidation efficiently.

Hamdi and Chainbi (2020) proposed the multi-weight strategy for consolidation, in which the constraints are also considered along with the resources in consolidation. Initially, the hosts were selected based on the resource consumption values and later were tagged with some weights based on the constraints associated with them. By coordinating the resource consumption values and constraints, the consolidation process was executed by them. Huang and Tsang (2014) worked on the M-convex model of consolidation to control the number of migrations and host shutdowns while consolidating. Apart from the main resource consumption values, host reconfigurations are also considered in this model to make the consolidation more energy-efficient. Although the former researches considered the other aspects along with resource consumption, there is still some gap appearing in managing the consolidation with multi objectives.

Materials and Methods

This section describes the proposed consolidation model MRCCM in detail. Before the discussion of the MRCCM, the data center architecture diagram layers with resource allocation, single resource-based allocation issues, and the need of considering multi resource-based consolidation will be explained in the form of the problem statement.

Data center Resource Allocation

Datacenter DC is constructed with a set (P) of Hosts (H), where $P = \{H_1, H_2 \ldots H_m\}$ and each host split into a set (Q) of virtual machines (V), where $Q = \{V_1, V_2 \ldots V_n\}$. In general, the hosts are equipped with K different types of resources R (i.e., CPU, RAM, BW, and HD) as $R = \{R_1, R_2 \ldots R_k\}$ and later these resources are partitioned and assigned to VMs. As shown in Fig. 2, on top of the PMs, the hypervisors will create the virtual layer with VMs for efficient utilization of resources. Each VM behaves as a virtual server and can process the client jobs on demand. According to the job requirement, some parts of the host resources are dissected from the host and have been allocated to the VM. Each host may contain one or more VMs, which can access and share the host resources among them.

Table-1 is presenting a set of frequently using MRCCM parameters with their detailed description for understanding. In the beginning, the Total Granted Resources (TGR) of a host H_i is marked as $TGR(H_i)$, from TGR (H_i) the resources are dissected and assigned with Q, where $Q \in H_i$ and the total utilization of the H_i is $TUR(H_i)$ is calculated as:

$$TUR(H_i) = \sum_{i=1}^{n} TUR(V_i) \rightarrow \quad (1)$$

In the case of multi-resource-based consolidation statistics the same Eq. (1) would be represented in the matrix format as follows:

DOI: 10.3844/jcssp.2022.665.673

$$TUR\left(H_i\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}=TUR\left(V_1\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}+TUR\left(V_2\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}+\dots+TUR\left(V_n\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}\rightarrow \quad (2)$$

The Total Utilization of the Resources value (TUR) of a host plays a vital role in deciding the need for consolidation. Once the consolidation process is confirmed, then the TUR will be deducted from TGR to estimate the TAR of a host H_i as follows:

$$\left(H_i\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}=TGR\left(H_i\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}-TUR\left(H_i\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}\rightarrow \quad (3)$$

Before the consolidation process begins, the best target hosts should be selected based on the available resources at the host. In general, the Total Available Resources (TAR) value of a host is compared against the migrating VM demanding resources to verify whether the host is suitable or not:

$$TDR\left(H_i\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}=TDR\left(j_1\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}+TDR\left(j_2\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}+\dots+TDR\left(j_n\right)\begin{bmatrix}R_1\\R_2\\\dots\\R_k\end{bmatrix}\rightarrow \quad (4)$$

To avoid aggressive overloading issues in the future, along with the host's current resource consumption value (TUR), the total user jobs of the most demanding (reserved) resources TDR (H_i) are also to be considered by calculating as shown in Eq. (4).

Single Resource-Based Consolidation Problem

As discussed above, most of the former research works followed the single resource (either CPU or RAM) based consolidation to avoid energy extravagance by pausing the underloaded host systems. To identify the under-loaded hosts, the host's current CPU utilization value will be compared against the CPU utilization lower threshold and the hosts which are utilizing the CPU less than the threshold are tagged as underloaded hosts and will be readied for consolidation. In this model, only one resource (i.e., CPU) is considered to determine the load on the host. In reality, the other resources (i.e., RAM, BW, HD, etc.) also have enough importance in processing the user tasks, hence it is recommended to consider the other resources as well in determining the underloaded hosts. This section illustrates the real-time use case, which taught the prominence of the other resources in consolidation.

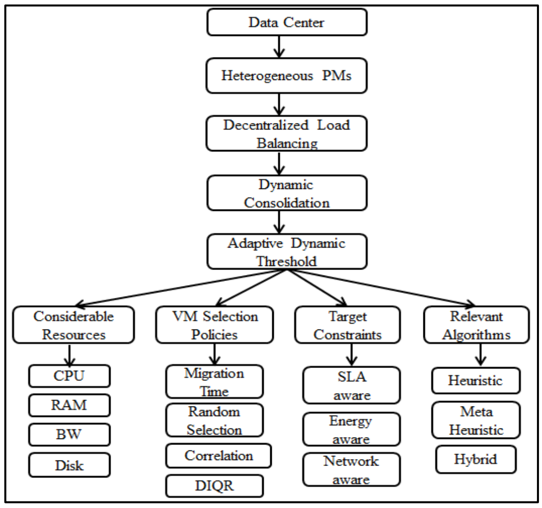

Fig. 1: Classification of the considerable objectives in consolidation

Before the use case discussion, a brief overlook of different types of jobs processed by the cloud datacenters was presented. In general, the end user jobs/tasks are designed to perform the CRUD operations at the data center-hosted application level. The majority of the user tasks are scheduled automatically on arrival, to respective VMs for immediate processing, but a few of them will wait in a queue for batch processing. Among the automated scheduling jobs, most of them needed CPU time to process and a good amount of Memory (RAM) to load the data for the process. According to Oracle Learning Cloud's report on automatically running jobs, a group of jobs needed slightly higher memory to load the job data for processing. Generating the video thumbnails after uploading, sending the analysis reports to the user, generation of the user profile with docs and images, video and audio relevant play and upload requests are the best examples, in which the memory and bandwidth are needed to be higher than the processing power.

For example, consider a running host, with lower threshold values CPU-38.5, RAM-42.1, and BW- 34.5%. The current resource utilization scores *(TUR* (H_i)) of the host is CPU-32.7, RAM-81.0, and BW-74.6%. According to the single resource (CPU) based consolidation model, this host H_i is underloaded (32.7<38.5) and is eligible for consolidation. Although the other resources like RAM and BW are utilized at their peak (higher threshold) level, this host *Hi* is considered for consolidation as the CPU consumption (32.7) is less than the lower threshold (38.5). This type of consolidation is called an inappropriate consolidation and it is most expensive because it needs another full host to allocate these high amounts of memory and network resources. Apart from that, it consumes more time for shifting the huge data, and this causes energy extravagance as it needs another new host to be shifted. In short, this consolidation is not considering the all effectible resources, selects inappropriate hosts, promotes aggressive migrations, and finally caused energy extravagance.

DOI: 10.3844/jcssp.2022.665.673

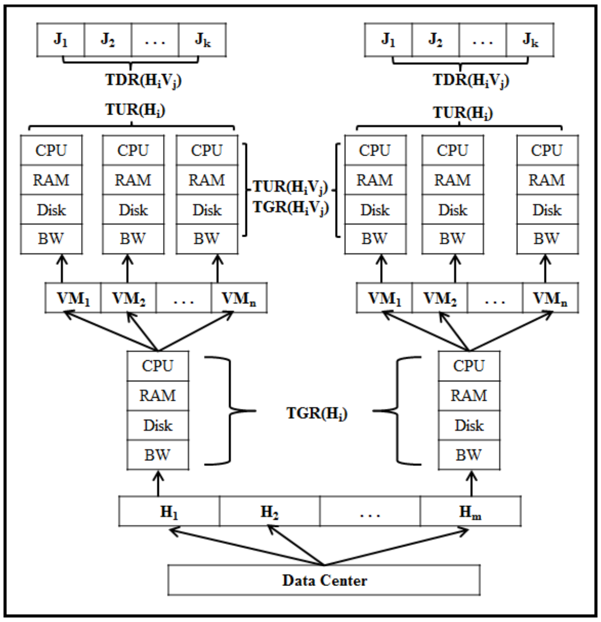

Fig. 2: Data center architecture diagram with layers and multiple resources

Hence it is recommended that, the multi resource-based consolidation for an efficient and reliable consolidation process, which considers the other prominent resources while selecting a host for the consolidation.

Multi Resources and Constraints Based Consolidation Model

By inspiring from the aforementioned single resource-based consolidation model issues, this study proposed the Multi Resources and Constraints-based Consolidation Model (MRCCM), which considers the multiple resources and resource constraints, while selecting the hosts for consolidation. The main contributions of this model are (i) Considering the effectible resources in consolidation, (ii) best target host selection for VM consolidation (iii) reduction of aggressive migrations, and (iv) mitigating the energy extravagance.

According to Patil and Patil (2020) proposed heuristic consolidation model, the data center-contained hosts are categorized as, the running costs under the execution pool, sleeping hosts under the sleep host pool, and ready-to-use hosts under the active host pool and shut downed hosts under inactive host pool. In contrast to the other models, the authors classified the hosts of a data center into these four different pools to control the energy extravagance and to satisfy the SLA. In their model, the running hosts are ordered based on their resource utilization, hence the detection of the underloaded or overloaded hosts becomes easy and completed in less time. By implementing this greedy consolidation model of the data center, a considerable amount of energy is saved and also controlled the aggressive migrations. But the main limitation observed in this model is the single resource-based consolidation. Our MRCCM follows a similar kind of data center architecture with four different pools but executes the consolidation based on multiple resources and constraints.

MRCCM consists of several steps in implementing the consolidation processes: Host-VM mapping, weights allocation, ordering of the hosts, consolidation triggering, underloaded host detection, target host selection, VM migration, and Host consolidation as shown in the MRCCM algorithm.

Algorithm-1: MRCCM

Input: Host set P, VM's set Q, Resource set R
Output: Ordered and Consolidated Host set Ṕ
Begin:
$\forall H_i \in P$, map $H_i \rightarrow Q_i$
|For each resource R_i,
| Calculate the $CUS(R_i)$ and $CAS(R_i)$
| Assign the weight $W(R_i)$ (e.q [7])
|end
| O(P) = OrderHostsByMultiResource(P)
| |Foreach H_i in Hostlist O(P), where ($H_i \in \delta_L$\{O(P)\}) do
| | |If(H_i is set for consolidation) then
| | | |Foreach V_j in $Q_i \rightarrow H_i$ do
| | | | Compare the TDR(V_j) with TAR(H_k)
| | | | |If(TDR(V_j) < TAR(H_k)) then
| | | | | Migrate V_j from H_i to H_k
| | | | |endif
| | | |end
| | |endif
| |end
| return the P as Ṕ
|**End:**

In MRCCM, the consolidation process starts with mapping the relations among the data center running hosts and their associated VMs. A running host H_i (where $H_i \in$ P), it's associated with each virtual machine V_j and the assigned job set J are mapped in a ternary relational model, to represent the relations and dependencies among them. At this moment, the mapping attributes are created to represent the associated mapping values and relations. Once the mapping of the elements is completed then the weights should be assigned further to decide the priority of each resource in the consolidation process. As per the former consolidation models by Park and Pai (2006), the CPU consumption is only calculated for consolidation, which is suffering from the above-mentioned single resource-based limitations. This research paper assigns the high priority value (weight) for the CPU and calculates the weights for the other resources also to build the multi-resource-based consolidation process. For this, the Cumulative Utilization Score (CUS) value of each resource across the running costs will be calculated. Later the same resource-related Cumulative Allocation Score (CAS) is also calculated and finally, the weight of the resource W(R_i) is determined according to the utilization rate of that resource as follows:

$$CUS(R_i) = \sum_{j=1}^{m} TUR(H_j)[R_i] \rightarrow \tag{5}$$

$$CAS(R_i) = \sum_{j=1}^{m} TAR(H_j)[R_i] \rightarrow \tag{6}$$

$$W(R_i) = \frac{CUS(R_i)}{CAS(R_i)} \rightarrow \tag{7}$$

It means the highly consuming resource will get a higher weight than the less consuming resources. In this way, all resource (except CPU) weights are calculated and assigned to users in the decision-making process of the consolidation. Like this, the MRCCM is considering the prominent and effectible resources in consolidation with respective weights. Based on the resource weights and consumption values, the running hosts will be categorized and ordered at regular intervals at the data center level. The total running hosts are divided into High (δ_H), Mid (δ_M), and Low (δ_L) workload categories based on resource utilization value. Optimal higher and lower threshold values are calculated for each resource and those threshold values are compared against the current utilization of respective resources to decide the workload category. The hosts with resource utilization > δ_H belong to a high category, utilization < δ_L belongs to the low category, and the rest of them with utilization between δ_H and δ_L are belongs to the mid category of utilization. As the consolidation is required for the under-loaded hosts, MRCCM concentrates only on the low utilization category of hosts, for which the consolidation is required. In this way, the MRCCM reduced the number of hosts at the data center to be searched to find the underloaded hosts for consolidation. Patil and Patil (2020) used only one resource (CPU), the categorization process happened based on CPU consumption against thresholds. But in MRCCM, multi-resource consumption-based categorization should be implemented, which is an NP-Hard problem. Algorithm-2 describes the proposed host categorization and ordering model based on multiple resource utilization values and weights.

Algorithm-2: Multi resource-based ordering of host set

```
OrderHostsByMultiResource(P)
Input: Host set P
Output: Ordered Host set O(P)
ResourceLowerThresholds δL(Rcpu), δL(Rram), δL(Rbw)
UnderLoadedHostSet A1, A2, B1, B2, C1, C2 and C3
OrderedHostSet O(P) = null
Begin:
# Comparison with Resource Thresholds foreach host Hi in P
  do
      if (TUR(Hi)[Rcpu] > δL(Rcpu) && TUR(Hi)[Rcpu] <
      δL(Rcpu)*1.25) then
          if (TUR(Hi)[Rram] < δL(Rram)*0.25 && TUR(Hi)[Rbw]
          < δL(Rbw)*0.25) then
            add Hi to A2
          else if (TUR(Hi)[Rram] < δL(Rram)*0.50 &&
          TUR(Hi)[Rbw] < δL(Rbw)*0.50) then
            add Hi to A1
          endif
      else if (TUR(Hi)[Rcpu] < δL(Rcpu) && TUR(Hi)[Rcpu] >
      δL(Rcpu)*0.75) than
          if (TUR(Hi)[Rram] < δL(Rram)*1.25 &&
          TUR(Hi)[Rbw] < δL(Rbw)*1.25) then
            add Hi to B2
          else if (TUR(Hi)[Rram] < δL(Rram)*1.50 &&
          TUR(Hi)[Rbw] < δL(Rbw)*1.50) then
            add Hi to B1
          endif
      else if (TUR(Hi)[Rcpu] > δL(Rcpu)*0.75) than
          if (TUR(Hi)[Rram] < δL(Rram) && TUR(Hi)[Rbw] <
          δL(Rbw)) then
            add Hi to C3
          else if (TUR(Hi)[Rram] < δL(Rram)*1.25 &&
          TUR(Hi)[Rbw] < δL(Rbw)*1.25) then
            add Hi to C2
          else if (TUR(Hi)[Rram] < δL(Rram)*1.5 &&
          TUR(Hi)[Rbw] < δL(Rbw)*1.5) then
            add Hi to C1
          endif
      endif
  end
add from A1 … C3 in order to O(P)
return O(P)
```

The MRCCM arranges the hosts in an order based on the multiple resources utilization value. This process helps to categorize the data center hosts into High, Mid, and Low ranges. For each resource R_i, the higher and lower utilization thresholds will be calculated according to our former research proposals by Beloglazov and Buyya (2010). To order the hosts for consolidation, the lower threshold values of various resources like $\delta_L(R_{cpu})$, $\delta_L(R_{ram})$, and $\delta_L(R_{bw})$ are calculated. Later each host-related total CPU utilization value $TUR(H_i)[R_{cpu}]$ is compared against the cpu lower utilization threshold $\delta_L(R_{cpu})$ value in various dimensions to categorize the hosts and place them in order as shown in algorithm-2. As the CPU is having a high priority than the other resources, MRCCM orders the hosts based on CPU utilization first. Later the hosts will be categorized based on the utilization value of the other resources like RAM and BW according to their weights. Finally, the categorized hosts will be added to the respective underloaded host set (i.e., A_1 … C_3). This process will continue iteratively, till all hosts are categorized to the target host category. Finally, all host sets are added to the ordered host set O(P), in order from A_1, A_2, B_1 … C_3. In this final ordered set, all from C_1 to C_3 are under-loaded hosts only, whereas the C_3 is consuming less quantity of resources than its preceded hosts. As the order of the hosts O(P) follows the descending order, the consolidation process will begin from the end of that order (i.e., from C_3). In this way, the least consuming (highly underloaded) hosts are consolidated before the other underloaded hosts. This process reduces the need for the consolidation of all underloaded hosts, in this way the number of consolidations and the aggressive migrations will be reduced by the MRCCM.

After the host order is completed, the consolidation process will begin to shift the VMs from underloaded hosts for load balancing. For this, among the under-loaded hosts, the least loaded hosts (C_3) related total resource consumption value *TDR*(*Hi*) is compared against the other hosts related total available resources value *TAR*(*Hi*), to determine the best target host to migrate the VMs from the under the loaded host. In this method, one least under-loaded host (C_3) VMs will be shifted to another under-loaded host (C_1). Instead of consolidating all under-loaded hosts (i.e., from C_1… C_3) with others (A_1 … B_2), MRCCM maximum prefers the consolidation among the under-loaded hosts (C_1 … C_3) in themselves. In this way, the number of consolidation requirements will be reduced along with migrations. Once the consolidation is completed, then the idle hosts will be assigned to either the sleeping pool or shutdown pool to save energy. In general, the service providers will schedule this consolidation process either at regular intervals or based on the data center resource shortage rates.

Experiments

To expand the horizons and to prove the efficiency of the proposed MRCCM, the experiments were conducted with real-time cloud-generated workload data. For this, the cloudsim toolkit by Calheiros *et al*. (2011) is used as the processing framework and the planet lab dataset is used for the real-time cloud resource consumption values. Cloudsim is the most popular simulation tool kit for cloud experiments and the planet lab data was collected from the real cloud data centers on scheduled days for conducting the experiments.

For MRCCM experiments, a total of 800 hosts of two different types (HP ProLiant-ML110G4 and ML110G5) were selected. ML110G4 is an equivalent range of Intel Xeon 3040 processor and ML110G5 is equivalent to Intel Xeon 3075. Both are single-core processors and are configured with different resource values for reliable comparison. By inspiring the standard Amazon EC_2 model, similar VM models were designed for these experiments. To strengthen the reliability of the experimental results, this study used the former proposed consolidation methods for experiments. Our experiments compared the MRCCM against various standard consolidation models (i.e., MRCCM, DIQR, IQR, MAD, LRR, and THR) using several properties like TEC, TNM, NH_Shut, ASLAV, HSM, and HSSD as shown in Table 2.

In each consolidation model, the Minimum Migration Time (MMT) has been considered for the target host selection to satisfy the Service Level Agreement (SLA) constraints. MRCCM obtained results are compared against our former research presented consolidation models are Dynamic Inter Quartile Range (DIQR) Inter Quartile Range (IQR), Mean Absolute Deviation (MAD), Local Regression Robust (LRR), and Static Threshold (THR). After the planet lab workload is executed by all these consolidation models, the results are compared using Total Energy Consumption (TEC), Total Number of Migrations (TNM), SLA violation percentage (SLA), Number of Hosts Shut down (NH_Shut), Average SLA Violation (ASLV), Host Selection Mean (HSM) and Host Selection Standard Deviation (HSSD) as shown in Table 2.

Results and Discussion

As our MRCCM main goal is controlling the energy extravagance, and reducing the number of migrations and consolidation, this study compared the experimental results with the respective metrics only. Figure 3 is presenting the MRCCM consumed total energy (i.e., 111.84 Kwh) in comparison with the other prominent consolidation models. Here the MRCCM saved energy from the 40-78 Kwh while processing the same workload, which is a considerable improvement in terms of energy consumption.

Reducing the number of consolidations and migrations is another research objective of this study. Table 2 describes that the number of host shutdowns (consolidations) is dramatically reduced with MRCCM, in which the MRCCM records 1853, whereas the DIQR, records 3244 and IQR records 5827 shutdowns. On another hand, Fig. 4 is presenting the comparison of the migrations that happened during the workload processing by various consolidation models. Among them, the MRCCM reported a less number of migrations (11570), when compared to the other models with a range from 23014 to 27632. These migration and consolidation statistics are showing that the MRCCM recorded a considerable reduction in the number of migrations and consolidations.

Table 1: Description of the parameters used with MRCCM

Params	Description
DC	Datacenter
H_i	i^{th} Host
V_i	i^{th} VM
TGR (H_i)	Total granted resources of i^{th} Host
TUR (H_i)	Total utilizing resources of i^{th} Host
TDR (H_i)	Total demanding resources of i^{th} Host
TAR (H_i)	Total available resources of i^{th} Host
TGR (V_i)	Total granted resources of i^{th} VM
TUR (V_i)	Total utilizing resources of i^{th} VM
TDR (j_i)	Total demanding resources of i^{th} Job

Table 2: Comparison of the MRCCM with former consolidation models

Consolidation model	TEC (Kwh)	TNM	SLA (%)	NH _Shut	ASLAV (%)	HSM (sec)	HSSD (sec)
MRCCM_MMT	111.8	11570	0.002	1853	9.55	0.007	0.004
DIQR_MMT	151.3	23014	0.002	3244	9.31	0.008	0.004
IQR_MMT	188.8	26476	0.003	5827	9.98	0.009	0.002
MAD_MMT	184.8	26292	0.003	5759	10.18	0.001	0.015
LRR_MMT	163.1	27632	0.004	5023	9.60	0.008	0.006
THR_MMT	177.3	24492	0.003	5477	10.34	0.006	0.003

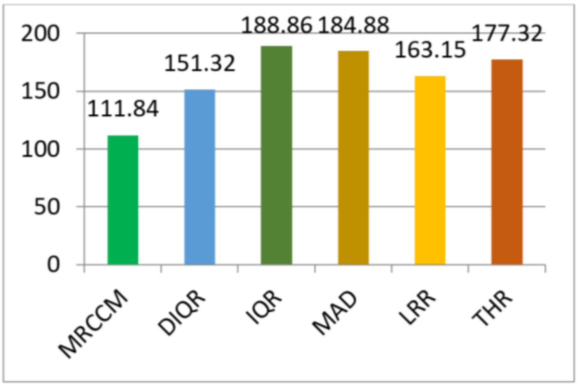

Fig. 3: Comparison of the total energy consumption (in Kwh) with MRCCM

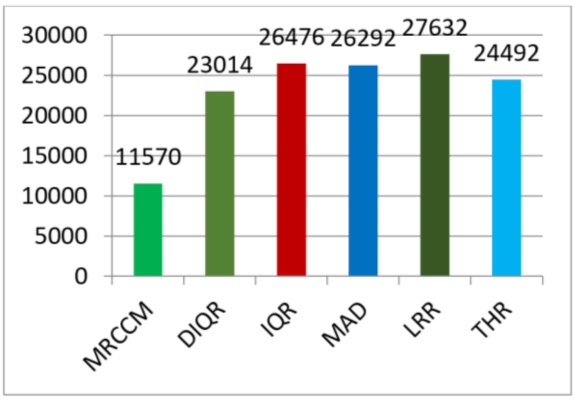

Fig. 4: Comparison of the total VM migrations with MRCCM

Conclusion

This study discussed the single resource-based consolidation limitations and the need for multi-resource-based consolidation in an elaborated manner. Although the former research works accepted the importance of the multi-resource-based consolidation, but not implemented in their research works as the multi-resource-based consolidation is an NP-Hard problem. To address the limitations of the single resource-based consolidation problem, the Multi Resource and Constraints-based Consolidation Model (MRCCM) was proposed. Along with the CPU, the MRCCM considers other resources like RAM and BW consumption also in the decision-making of the consolidation. Experiments on multi-threshold-based host categorization and multi-resource constraints-based ordering helped in reducing the number of consolidations and migrations along with the energy extravagance.

In the future, MRCCM experiments are planned to implement novel consolidation algorithms and constraints to satisfy the multiple resources at a time.

Acknowledgment

I would like to express my sincere gratitude to my guide Rekha Patil for her immense support throughout this research work. The support of CSE department staff of my university in infrastructure setup and experiments evaluation is appreciable.

Author’s Contributions

Archana Patil: Design of the work, data collection, data analysis, drafting the article and revision of the article.

Rekha Patil: Guiding to write the paper, drafting the article, revision of the article and supporting to write the paper.

Ethics

The substance of this manuscript is authors original research work and doesn’t contain any plagiarized or copy righted content in any form. All authors read and approved that, this manuscript doesn’t violated any publication ethics of the journal.

References

Ahmad, R. W., Gani, A., Hamid, S. H. A., Shiraz, M., Yousafzai, A., & Xia, F. (2015). A survey on virtual machine migration and server consolidation frameworks for cloud data centers. *Journal of Network and Computer Applications*, 52, 11-25, doi.org/10.1016/j.jnca.2015.02.002

Beloglazov, A., & Buyya, R. (2010). Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers. *MGC@ Middleware*, 4(10.1145), 1890799-803.

Beloglazov, A., & Buyya, R. (2012). Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. *Concurrency and Computation*: *Practice and Experience*, 24(13), 1397-1420. doi.org/10.1002/cpe.1867

Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A., & Buyya, R. (2011). Cloud Sim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software*: *Practice and Experience*, 41(1), 23-50. doi.org/10.1002/spe.995

Ferdaus, M. H., Murshed, M., Calheiros, R. N., & Buyya, R. (2014, August). Virtual machine consolidation in cloud data centers using ACO metaheuristic. In *European Conference on Parallel Processing* (pp. 306-317). Springer, Cham. doi.org/10.1007/978-3-319-09873-9_26

Gangadharan, G. R., & Murugesan, S. (2012). Harnessing Green IT: *Principles and practices*. doi.org/10.1002/9781118305393

Hamdi, N., & Chainbi, W. (2020, May). A multi-weight strategy for container consolidation. In 2020 *IEEE 4th International Conference on Fog and Edge Computing (ICFEC)* (pp. 11-18). *IEEE.* doi.org/10.1109/ICFEC50348.2020.00009

Huang, Z., & Tsang, D. H. (2014). M-convex VM consolidation: Towards a better VM workload consolidation. *IEEE Transactions on Cloud Computing*, 4(4), 415-428. doi.org/10.1109/TCC.2014.2369423

Khan, M. A., Paplinski, A., Khan, A. M., Murshed, M., & Buyya, R. (2018). Dynamic virtual machine consolidation algorithms for energy-efficient cloud resource management: *A review. Sustainable Cloud and Energy Services*, 135-165. doi.org/10.1007/978-3-319-62238-5_6

Li, H., Zhu, G., Cui, C., Tang, H., Dou, Y., & He, C. (2016). Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing. *Computing*, 98(3), 303-317. doi.org/10.1007/s00607-015-0467-4

Minas, L., & Ellison, B. (2009). Energy efficiency for information technology: How to reduce power consumption in servers and data centers. *Intel press*. doi.org/10.5555/2826078

Park, K., & Pai, V. S. (2006). CoMon: A mostly-scalable monitoring system for Planet lab. *ACM SIGOPS Operating Systems Review*, 40(1), 65-74. doi.org/10.1145/1113361.1113374

Patel, S., & Makwana, R. M. (2016). Optimized energy efficient virtual machine placement algorithm and techniques for cloud data centers. *Journal Computer Science*, 12(9),448-454. doi.org/10.3844/jcssp.2016.448.454

Patil, A., & Patil, R. (2020, December). An optimal heuristic model for greedy and efficient consolidation process in cloud datacenters. In 2020 *3rd International Conference on Intelligent Sustainable Systems* (*ICISS*) (pp. 1308-1313). IEEE. doi.org/10.1109/ICISS49785.2020.9316026

Sabahi, F. (2012). Secure virtualization for cloud environment using hypervisor-based technology. *International Journal of Machine Learning and Computing*, *2(1)*, *39*. doi.org/org/papers/87-A888.pdf

Saranya, S. S. & Fatima, N. S. (2021). Efficient handling of medical data classification in the cloud-edge network using an optimization algorithm. *Journal of Computer Science*, 17(11), 1116-1127. doi.org/10.3844/jcssp.2021.1116.1127

Uhlig, R., Neiger, G., Rodgers, D., Santoni, A. L., Martins, F. C. Anderson, A. V., & Smith, L. (2005). Intel virtualization technology. *Computer*, 38(5), 48-56. doi.org/10.1109/MC.2005.

Xing, Y., & Zhan, Y. (2012). Virtualization and cloud computing. *In Future Wireless Networks and Information Systems* (pp. 305-312). Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-27323-0_39

Yu, L., Chen, L., Cai, Z., Shen, H., Liang, Y., & Pan, Y. (2020). Stochastic load balancing for virtual resource management in data centers. *IEEE Transactions on Cloud Computing*, 8(2), 459-472. doi.org/10.1109/TCC.2016.2525984

Zhang, X., Shae, Z. Y., Zheng, S., & Jamjoom, H. (2012, April). Virtual machine migration in an over-committed cloud. In 2012 *IEEE Network Operations and Management Symposium* (pp. 196-203). IEEE. doi.org/10.1109/NOMS.2012.6211899