Migrating Virtual Machine in Heterogen and Homogen Cloud Data Center with Efficient Electricity Consumption Using Fuzzy C-Means

Guruh Fajar Shidik¹* Edi Noersasonko¹ Adhitya Nugraha¹ Oki Setiono¹

¹Computer Science Department, Universitas Dian Nuswantoro, Indonesia
* Email: guruh.fajar@research.dinus.ac.id

Abstract: Dynamics Virtual Machine management (VM) is a solution for reducing the huge resource and energy usage problems in the cloud data center. By managing the VM usage, it can decrease the operating cost and mitigate the spread of carbon dioxide (CO2) into the environment. This research presents the used of Fuzzy C-Means clustering to select VM in an overloaded host and distributes it into another available host. The proposed method was implemented in workloads dataset which provided by CloudSim PlanetLab. From evaluation process, the proposed method can produce better performance in reducing energy consumption, especially using four clusters in Heterogen workload comparing with other VM selection techniques such as Random Choice and Minimum Time Migration with reduction percentages are 9.34% and 3.89% respectively. This evaluation result denoted that our proposed method has better performance in reducing EC in the cloud data center.

Keywords: Cloud data center, Dynamic VM integration, Energy efficiency, Fuzzy C-Means, VM selection.

1. Introduction

Cloud computing is one of the forms of technological progress in the field of Green IT. In the IoT era, cloud computing became one of the best integrations for companies especially in Information Technology (IT) to store their information into the data center. However, cloud data center use huge amounts of electricity resource which raise the operating costs and increasing the spread of CO2 into the environment [1].

In addition, cloud computing technology has an impact on large emissions of CO2, ranges from 2% of global emissions [2]. Thus, the cloud computing technology requires energy efficiency management to reduce these impacts and ensure that computing resources can be used efficiently to continue to run the application as well as minimizing energy usage while maintaining the necessary quality of service (QoS) [3].

The energy-aware mechanism in the cloud data center is also an important part to solve the issues in reducing energy consumption [4]. To enhance the energy efficiency in the data center, the Dynamic VM integration is considered as one of effective management solution [5]. Dynamic VM integration has four tasks of decision making [6, 7]: (1) determine when the hosts are redundant (Host Overload Detection), based on this condition, direct migration is involve to move the VM(s) from host when its overloaded; (2) decide when the hosts are underloaded, at this position, the host is available to switch into sleep mode, when this host’s VM(s) is migrated to another host; (3) Select the VM(s) which will migrated from the overfilled host (VM Selection); and (4) select the host location for the assignment of migrating VM(s) (VM Placement).

Moreover, lots of previous research in this area that discussed dynamic VM consolidation energy efficiency have been done such as in [5-10]. Where some research only focus on decreasing energy consumption in VM migration [11], or allocation [12]. Other previous work tried to emerging the statistical method in VM selection, such as fuzzy Markov model [13, 14], time-series forecasting approach [15] or adaptive three-threshold [16]. Moreover, other
studies [8, 17, 18] focused on the VM selection utilized clustering techniques.

In previous research [18] which utilized K-means clustering combined with normal Markov Model was proposed. This research applied a proposed combination to optimize the VM selection due to the used of power resource in the data center. That proposed method could increase the efficiency in the cloud data center up to 6.8%. Another study proposed by Masoumzadeh [19] was using Fuzzy-Q Learning to make an optimal determination in migrating the VM(s). The result of this study provides better results in the performance of energy trade-offs in the cloud data center compared with another algorithm. It does show that Fuzzy technique can be applied in optimal decision making in selecting VM(s) to migrate.

Based on the research above, this study will be focused on VM Selection that will affect the performance when the host is overloaded. This research proposed the used of Fuzzy C-Means clustering technique to analyze the pattern of the VMs tendency in heterogeneous and homogenous condition since there is no research which implemented the Fuzzy C-Mean as the clustering algorithm in VMs selection. Cluster analysis defined as the process of dividing the VMs into a set of several groups based on their similarity distance [20].

Moreover, FCM has ever been used in a different area of study proposed by Ali Feizollah et al [21], in this paper, they introduced the use of Fuzzy C-Means clustering in Android malware detection. They chose 800 malware samples, and 100 clean applications and collected generated network traffic. The features were generated from the network traffic, and after that used it in FCM clustering. The results have shown that this proposed algorithm was able to cluster the data into two groups of clean and malicious data. Furthermore, the proposed method was validated by comparing the result to a labeled dataset, which showed 13% of discrepancy in results.

FCM is considered as a VM selection method in this research because it has a more natural approach than the classical or statistical clustering. This research will examine the implementation of Fuzzy C-Means Clustering in selecting the destination host to move VMs from the overfilled host. The proposed algorithm will be analyzed and compared to other existing methods in previous studies such as Random Choice (RC) and Minimum Time Migration (MMT) [13].

2. Proposed FCM VM selection

In this section, the implementation of FCM clustering in VM selection process described as follow:

1. Determine the attributes for clustering. In this case, the required attributes are RAM and MIPS. For the size of data, it will match the amount of the VM in one host on each tested workload.

2. Define the parameters used in Fuzzy C-Means.
   - Number of Clusters : 3 or 4 or 5
   - Weighting : 2
   - Maximum Iteration : 50
   - Error Rate : 0.01
   - Objective Function : 0
   - Early Iterations : 1

The number of the cluster could be more than 2. In this experiment, we have used 3, 4, or 5 clusters to find out the optimal variance of the cluster member that suitable in this case. Moreover, the weighting value is set to 2 because its value should be more than 1 (Weight > 1) [22]. Then, the maximum iteration is set to 50, which means that the process will be running up to 50 times. However, if the value of error rate or objective function is fulfilled even the maximum iteration is not reached yet, the clustering process will be stopped. Early iteration was set into 1 means that the iteration was iterated from first index.

3. Once the parameters are defined, raised the partition matrix. This partition matrix is used to determine the data that will be entered into clusters according to the largest value. The line numbers will match the amount of available data, and the columns are to adjust the number of clusters. Below is an example of the partition matrix with three clusters. Noteworthy is that each line number of each value should be worth one.

4. Calculate the cluster centers for each cluster with the formula.

\[ V_k = \frac{\sum_{i=1}^{n}(\mu_{ik})^w X_i}{\sum_{i=1}^{n}(\mu_{ik})^w} \]  

Table 1. Sample of matrix partition with 3 cluster

<table>
<thead>
<tr>
<th></th>
<th>k1</th>
<th>k2</th>
<th>k3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>0.1</td>
<td>0.2</td>
<td>0.7</td>
</tr>
</tbody>
</table>
In Eq. (1), $V_k$ represent the center cluster (centroid) of the cluster $k$. Then, $\mu_{ik}$ is the data partitioning (the matrix $\mu$) on the $i^{th}$ data and $k$ cluster. $X_i$ is the data (on the matrix $\mu$) on all clusters. Moreover, $W$ represent the weight or hyper-parameter, this value should fulfill $W > 1.0$. The bigger $W$ value which used in FCM, the fuzzier the cluster will be at the end of the process [22].

5. Calculate the objective value using Eq. (2) as shown below:

$$P_n = \sum_{k=1}^{n} \sum_{i=1}^{c} ((\mu_{ik})^w (d_{ik})^2)$$

(2)

From the equation above, can be described that $d_{ik}$ is distance measurement between data $i$ to the cluster center $k$, where $d_{ik}$ was used Euclidean distance measurement. Then, $w$ is weighting or hyper-parameter, and $P_n$ is an objective value at $n$ iteration.

6. Then in the final stage, the Fuzzy C-Means process is done by improving the partition matrix ($\mu$) that has been used in $P_i$. The steps are taken by improving the membership degree of each data in each cluster in the matrix partition. With the formula as follows:

$$\mu_{ik} = \left[ \sum_{j=1}^{c} \left( \frac{d_{ik}}{d_{jk}} \right) \right]^{-1}$$

(3)

$$d_{ik} = d(x_k - v_i) = \left[ \sum_{j=1}^{m} (x_{kj} - v_{ij}) \right]^{1/2}$$

(4)

From Eq. (3) and Eq. (4) can be known that:

- $\mu_{ik}$ : Data partitioning $i$ (the matrix $\mu$) on the cluster centers $k$.
- $d_{ik}$ : Euclidean distance from data $i$ to cluster center $k$.
- $W$ : Weight or hyper-parameter for fuzzier level determination.
- $x_{kj}$ : representation of the data (on the matrix $\mu$) on all attributes $j$ and cluster $k$.

7. The above process will continue repeatedly, after the difference between objective result $i$ and $i - 1$ functions is smaller than the error rate, or the number of iterations exceeds the maximum iteration value. The partition matrix will continue to be fixed until the stop condition is reached and the last matrix partition is used as the final result.

8. After the iteration stop or convergent by FCM, we will select the VM based on the largest index value of the matrix in each cluster. Based on the clustered index ($c_i$) on each VM, the VM will be selected from an overloaded host with the condition of the cluster at the following rules below:

$$x = \begin{cases} 
\{ x \in \min(c_i) \leftrightarrow c_i \neq \emptyset, \max(x) \\
\{ x \in \min(c_i+1) \leftrightarrow c_i = \emptyset, \max(x) \leftrightarrow c_{i+1} \neq \emptyset
\end{cases}$$

(5)

From the conditional equation in Eq. (5), can be known that $x$ is the VM instance which would be transferred from the host that overloaded. The first condition is to select the maximum VM instance ($\max(x)$) from minimum cluster index $c_i$, if the cluster index $c_i$ is not an empty set cluster ($c_i \neq \emptyset$). However, if the clustered index $c_i$ is empty set ($c_i = \emptyset$), then the second condition will select the maximum VM from the next minimum cluster index $c_{i+1}$ if only if the index $c_{i+1}$ has VM instance in its cluster. For more deep explanation of this phase, below is the implementation sample of selection VM after clustered with FCM in selecting VMs.

Suppose there exist clustered VM-ID which have 3 clusters as shown in Fig. 1(a). Each cluster usually has VM-ID as follows:

- Member of the first cluster : 11, 8, 5, 1.
- Member of the second cluster : 12, 10, 7, 3.
- Member of the third cluster : 9, 6, 4, 2.

These sample data above had been sorted descending based on its VM index. This sample case was fulfilled the first condition in Eq. (4) which is $c_i \neq \emptyset$ where there is no empty cluster. The first selected VM is a VM on a first cluster and VM-ID on the top position, in this case, is VM-ID 11. So that VM with ID 11 will be migrated. If the host is overfilled, then the next process is to move into the second cluster and choose VM-ID with the highest index which is VM-ID 12, then migrate that VM from the host. If the overload is still detected, the next step is selecting an existing VM on a third cluster with the top position which is VM-ID 9, then migrate it. These steps will be repeated until the host is not overloaded.
In the other condition that appears unfulfilled cluster member that shown in Fig.1(b), it should be used the second condition of Eq. (4) ($c_i = \emptyset$).

- Member of first cluster: $\rightarrow$ (empty cluster)
- Member of the second cluster: 8, 7, 3, 1
- Member of the third cluster: 9, 6, 4, 2.

In the above data (Fig.1(b)), there is an empty cluster which appears in the first cluster ($C_1$). This sample was represented as the second condition where empty clusters have existed. In this case, if the selection process of the VM is executed, clusters that do not have a member of the VM will be skipped to the next cluster. For instance, when a VM on third ($C_3$) cluster was selected (VM-ID 9), then this VM will be migrated. If the host is still overfilled, then the VM selection will directly select the VM from second ($C_2$) cluster, since the first cluster is empty / does not have a member.

3. Experiment and evaluation

3.1 Workload dataset

In Dynamic VM Consolidation, VM should be transferred from the overfilled host which has CPU workload. In this study, we performed experiments using workload dataset from PlanetLab [23]. The available data are gathered from daily workload CPU VM in the data center, on 03-03-2011 up to date 25-03-2011. Besides that, 800 hosts will be set up to be simulated from the workload VM. There are 5 different workloads which used to test the proposed model there are:

- Workload 2011-03-03 contains 1052 CPU VM
- Workload 2011-03-06 contains 898 CPU VM
- Workload 2011-03-09 contains 1061 CPU VM
- Workload 2011-03-22 contains 1516 CPU VM
- Workload 2011-03-25 contains 1078 CPU VM

3.2 VM instance category

The workload in each VM, required several specification details of VM instance while running at the heterogen and homogen condition. The VM instances that emulate workload data are set up from the standard Amazon EC2. VM samples that used in this study were heterogen with four different characteristics such as, High-level CPU (0.87 GB, 2500 MIPS), Medium level CPU (1.7 GB, 2000 MIPS) Small level CPU (1.7 GB, 1000 MIPS) and Micro level CPU (613 MB, 500 MIPS). Then, homogen VM instances characteristics using High-level CPU (0.87 GB, 2500 MIPS).

3.3 Evaluation

The proposed method was evaluated using SLATAH, Energy Consumption (EC), and PDM as shown in Fig. 2. The workloads CPU will be used similarly in both homogen VM and heterogen VM instances.
3.4 Requirement

This research was modelled and simulated using CloudSim tool to create Data Center simulation. The simulation tools were used 800 samples of a homogen physical host with HP ProLiant ML110 G5 Xeon 3075 series. The physical hosts have specification such as 4.096 RAM, 1Gbit/s, 2.660 MIPS, 2 PES, and 1000 GB Storage.

4. Result and discussion

The proposed method result of this study was obtained using CloudSim as a tool for simulating data centers. In this evaluation process, the heterogeneous and homogen VM(s) were used as cluster dataset. The results are assessed by comparing the evaluation result with other selection policies which are RC, and MMT. These selection algorithms have been used as comparison with [9].

4.1 Energy consumption (EC)

4.1.1. Heterogen

The results of the experiment in heterogen VM can be seen in Table 2, where workload 2011-03-03, 2011-03-06, 2011-03-09, 2011-03-22 and 2011-03-25 shown the decrement in the energy consumption (EC) value when using Fuzzy C-Means. The most efficient Energy Consumption in heterogeneous were using the Fuzzy C-Means method is when VM Selection uses 3 Clusters and 4 Clusters. With the average Energy Consumption of 153,86 kWh and 153,162 kWh. While the MMT algorithm as a comparison algorithm has the worst energy consumption results in each test of 5 workloads. The proposed method looks promising that achieve energy consumption reduction around 3.89% until 9.34% comparing with other VM selection techniques such as Minimum Time Migration and Random Choice.

4.1.2. Homogen

Based on the results of energy consumption testing in homogen VM which described in Table 3, it has shown the same results between the comparison method (MMT) and proposed FCM with 3, 4 and 5 clusters. However, the Table 3 shows that our proposed method is outperforming the RS method in energy consumption testing, with a proposed average result of 226.99 kWh compared to the RS method with an average of 229.43 kWh. The performance of the proposed method in reducing energy consumption almost similar with 1.07% higher than RS technique.

<table>
<thead>
<tr>
<th>Method</th>
<th>20110303</th>
<th>20110306</th>
<th>20110309</th>
<th>20110322</th>
<th>20110325</th>
<th>Average Consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM with 3 Clusters</td>
<td>163.83</td>
<td>124.88</td>
<td>146.74</td>
<td>179.33</td>
<td>154.52</td>
<td>153.86</td>
</tr>
<tr>
<td>FCM with 4 Clusters</td>
<td>164.97</td>
<td>124.75</td>
<td>144.48</td>
<td>178.70</td>
<td>152.91</td>
<td>153.16</td>
</tr>
<tr>
<td>FCM with 5 Clusters</td>
<td>165.18</td>
<td>125.81</td>
<td>146.29</td>
<td>179.58</td>
<td>154.05</td>
<td>154.18</td>
</tr>
<tr>
<td>MMT</td>
<td>179.53</td>
<td>134.94</td>
<td>157.44</td>
<td>199.69</td>
<td>173.12</td>
<td>168.94</td>
</tr>
<tr>
<td>RS</td>
<td>166.79</td>
<td>127.65</td>
<td>148.37</td>
<td>189.69</td>
<td>164.31</td>
<td>159.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>20110303</th>
<th>20110306</th>
<th>20110309</th>
<th>20110322</th>
<th>20110325</th>
<th>Average Consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM with 3 Clusters</td>
<td>250.61</td>
<td>189.40</td>
<td>214.03</td>
<td>260.89</td>
<td>220.02</td>
<td>226.99</td>
</tr>
<tr>
<td>FCM with 4 Clusters</td>
<td>250.61</td>
<td>189.40</td>
<td>214.03</td>
<td>260.89</td>
<td>220.02</td>
<td>226.99</td>
</tr>
<tr>
<td>FCM with 5 Clusters</td>
<td>250.61</td>
<td>189.40</td>
<td>214.03</td>
<td>260.89</td>
<td>220.02</td>
<td>226.99</td>
</tr>
<tr>
<td>MMT</td>
<td>250.61</td>
<td>189.40</td>
<td>214.03</td>
<td>260.89</td>
<td>220.02</td>
<td>226.99</td>
</tr>
<tr>
<td>RS</td>
<td>256.17</td>
<td>192.24</td>
<td>215.25</td>
<td>260.49</td>
<td>223.02</td>
<td>229.434</td>
</tr>
</tbody>
</table>
4.2 PDM

PDM or Performance degradation due to Migration describe the performance of VM which affected by VM migration. The lower of PDM percentage value will show the better performance of the Distributed Dynamic VM Consolidation SLA on the Cloud Data Centre [9].

4.2.1. Heterogen

From Fig. 3, the performance of PDM is decreased which caused by VM Migration.

The result can be seen in Fig 3, where using the Fuzzy C Means method with 3, 4, and 5 clusters increase the percentage of PDM compared to the MMT method. The percentage of each FCM clusters (3 clusters, 4 clusters, and 5 clusters) are 0,14%; 0,15%; 0,14% respectively.

However, when compared the 4 clusters with other clusters, the Fuzzy C-Means method with 3 and 5 Clusters experienced an average decline of 0,01%.

4.2.2. Homogen

Fig. 4 shown the results of the evaluation with PDM in homogen workload. Based on this figure, it can be seen that the FCM, MMT, and RS methods have the same average yield of 0,13%.

4.3 SLATAH

SLATAH stand for SLA Violation time per active host which represent the average time when active host used 100% CPU. This measurement can describe the distribution performance of VM in the cloud data center.

4.3.1. Heterogen

The lower the SLATAH percentage value, the better Distributed Dynamic VM integration on the cloud data center [9].

In heterogen VM dataset, the SLATAH percentage value when using FCM in VM Selection is higher when compared to the existing MMT and RS methods as shown in Fig.5. When using FCM in VM Selection using 3, 4 and 5 clusters the value has an average percentage 7,70%; 7,81%; 7,68% respectively.

The results of the percentage of the test value are higher than the MMT and RS methods, which are 5,79% and 7,59%.

4.3.2. Homogen

SLATAH evaluation results when using homogen in Fig. 6, shows the same result between FCM and MMT methods. However, the results have better results than the RS method. FCM and MMT have an average of 4,89%. Then the RS method has an average of 4,91%.
For the future works, the quality of service during migration VM will be improved to reduce the computation process with the extended rules-based.

Acknowledgments

This research has been funded by “Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jendral Penguatan Riset, Teknologi, dan Pendidikan Tinggi, Indonesia” with grant 026/L6/SP2H.1/PENELITIAN/2019

References

[10] A. Ashraf and I. Porres, “Multi-objective dynamic virtual machine consolidation in the


