Journal of Engineering Science and Technology
Vol. 16, No. 6 (2021) 5013 - 5026

ENERGY EFFICIENCY IN CLOUD DATA CENTER THROUGH UNSUPERVISED RULE-BASED VM SELECTION METHODS

GURUH FAJAR SHIDIK[1,*], ADHITYA NUGRAHA[1], PULUNG NURTANTIO ANDONO[1], EDI NOERSASONGKO[1], EDI JAYA KUSUMA[2]

[1]Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
[2]Faculty of Health Science, Universitas Dian Nuswantoro, Indonesia
*Corresponding Author Email: guruh.fajar@research.dinus.ac.id

Abstract

This research focuses on improving the Virtual Machine Selection mechanism in Dynamic virtualization consolidation to achieve robust energy-efficient consumption. The proposed unsupervised methods: K-Means or Fuzzy C-Mean (FCM) combined with rule-based MAX function. Both proposed methods were simulated, compared, and evaluated using the public workload dataset. The evaluation of this research measures the energy consumption in kilowatt-hour using CloudSim Environment. The simulation result shows that the highest decrement of energy consumption achieved by K-Means rule-based in 5 clusters until 16.51 kWh that followed by FCM rule based until 15.7 kWh compared with statistical RC and MMT VM selection methods. The statistical Friedman Test showed the improvement energy efficiency of both unsupervised rule-based with 0.0105 of the p-value. This result indicated a significant difference in energy consumption among all tested methods. Moreover, K-Means Rule-Based was able to be the first rank in the Friedman Test.

Keywords: Cloud computing, Cloud data center, Dynamic VM consolidation, Energy efficient, FCM, K-means, Rule-based, VM selection.

1. Introduction

The high demand for massive computation in all aspects encourages cloud computing, especially in Information and Communication Technology (ICT) based companies. The integration of cloud computing in ICT-based companies became one of the best solutions to storing its data and information centralized (cloud data center). For almost two decades, the cloud data center's energy efficiency issues have been discussed [1, 2]. This issue was raised due to cloud computing utilization consumes many energy resources, which causes the operation cost to increase. Moreover, cloud computing also responsible for carbon footprints, about 2% of the total global emission in the last two decades [3].

Due to this problem, the cloud data center requires energy management to reduce extravagant energy consumption, ensure the applications are running continuously, and maintain service quality (QoS). Dynamic VM Consolidation has become one mechanism to conserve energy consumption at the cloud data center [4]. There are four tasks of Dynamic VM Consolidation. The first is to determine or detect an overloaded host. If this condition is passed, the physical machine will migrate some virtual machines from an overloaded host. The second task is identifying the host in underload condition and move the leftover underload host VMs to other hosts and switch the underload host into sleep mode. The third task is selecting which VMs should be moved from an overloaded host. The last task is determining the location (host) to store the migrated VMs.

This research focuses on VM selection problems due to the complexity of selecting VM to increase energy efficiency in cloud data center. There are several types of VM selection methods proposed by previous researches such as Minimum Migration Time (MMT), Maximum Correlation (MC), Random Choice (RC), Fuzzy Q-Learning (FQL), and Constant Position Selection Policy (CPS) [4-8].

This study used unsupervised method that combine with rule-based method to improve the VM Selection mechanism in dynamic virtualization consolidation. The aim of the proposed method is to achieve robust energy-efficient consumption in cloud data center better than statistical VM Selection. The unsupervised method, also known as the clustering method, was chosen because of its capability to learn and produce natural and accurate cluster results [9]. Unsupervised clustering methods can be separated into two categories, such as partition hard clustering [10, 11] and fuzzy clustering [12, 13].

Therefore, this study will compare these two clustering algorithms, which are K-Means and Fuzzy C-Mean (FCM) clustering. The K-Means has been chose due to the simple unsupervised clustering procedures [14] and often used to solve many clustering problems [15-18]. Then, the FCM method has been chose due to a complex uniform effect that have good performance, and the fuzzifier parameter tends to affect the unsupervised clustering result [19]. Each proposed unsupervised methods that combine rule-based with MAX function, will be compared with popular and robust statistical VM selection methods are MMT and RC [20]. Both proposed methods and comparison methods have been simulated and evaluated using PlanetLab public workload dataset [21].

2. Related Research

A study regarding cloud computing has been done in several types of research [4, 6, 22-24]. Several studies concerning energy consumption have been proposed, such as VM migration [25] or VM allocation [26]. Another research conducted by Beloglazov and Buyya [5] proposed the adaptive heuristics method to improve the energy-efficiency in a cloud data center, where the method was focusing on allocating the VM in dynamic ways. Moreover, in this research, VM selection was conducted using Minimum Migration Time (MMT), Maximum Correlation (MC), and Random Choice (RC). MMT needs minimum time to move the VM, where minimum time can be calculated by dividing the RAM with available bandwidth. In contrast, MC will select the VM data to depend on the correlation between resources and an overloaded host probability. Moreover, RC will move the VM based on the distribution of the random variable. Then, MAD was applied to detect an overloaded host. The proposed method gives better energy efficiency compared with DVFS or non-power aware mechanism from the evaluation result.

The previous work [8] was used Constant Position VM Selection (CPS) to select which VM has to move from an overloaded host. This proposed method is working after overloaded hosts are identified. CFS provides better results from the evaluation result, especially in energy-efficiency and SLA, compared to other VM selection techniques that have been previously proposed (MMT, RC, and MC).

Based on the information above, it can be seen that many kinds of research have been proposed the VM selection method using a statistical approach to reduce the energy consumption in cloud computing. Masoumzadeh [7], proposes Fuzzy Q-Learning (FQL) in VMs selection. This research was implemented fuzzy logic to identify the overloaded physical machine in dynamic VM consolidation. The evaluation shows that FQL has achieved promising results regarding energy-efficiency in the cloud data center. However, this research focus on the static cluster selection depends on the fuzzification rules.

Based on previous works [22, 27, 28], unsupervised method such as K-Means or FCM have been implemented as a VM selection method. The implementation of each research was focused only on the number of clusters that suitable for VM selection. Based on the evaluation result, K-Means or FCM clustering showed promising energy-efficient results in the cloud data center. However, these research not directly compare the ability performance and robustness in same condition dataset, between K-Means and FCM that enhanced by Rule Based with MAX function.

Therefore, based on the related research above, this research proposed a continuing study that evaluates the performance of unsupervised K-Means and FCM clustering that has been improve by Rule Based with MAX function, in several cluster numbers for the VM selection process in cloud data center. Both proposed methods will be compared to conventional and robust statistical VM selection, such as MMT and RC, regarding the significant energy-efficiency improvement in the same condition.

3. Proposed Unsupervised Rule-Based VM Selection

The proposed unsupervised rule-based VM selection as seen in dash line Fig. 1 is part of Dynamic VM Consolidation [5]. The method works for minimizing the energy consumption in the cloud data center by selecting a virtual machine that should be moved away from the overloaded physical machine.

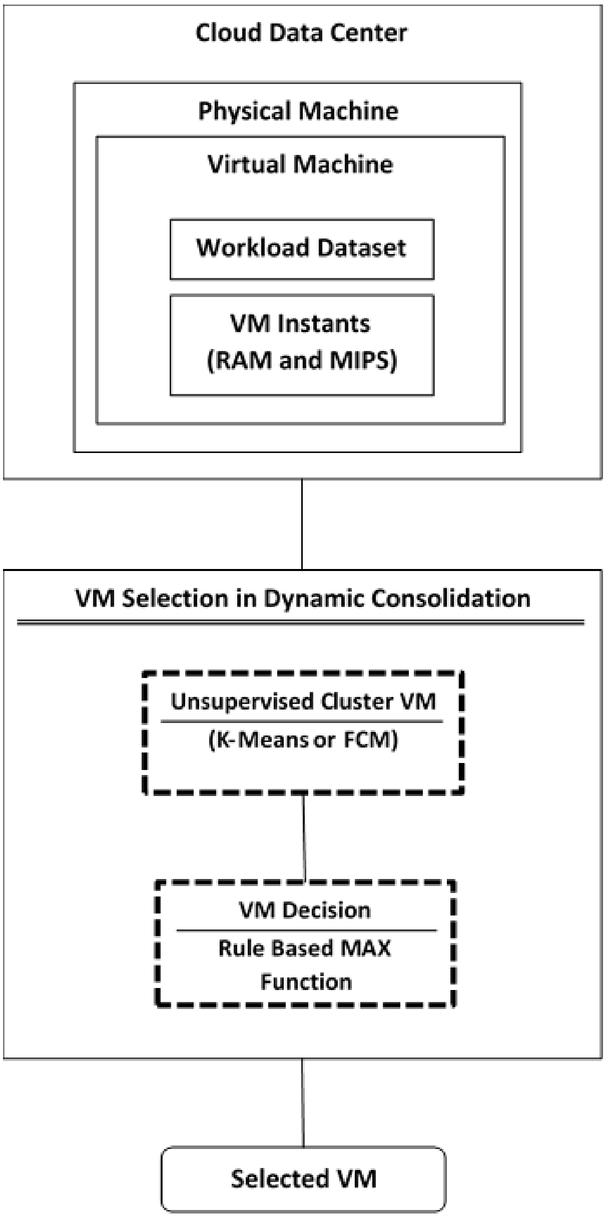

Fig. 1. Proposed unsupervised VM selection method.

The unsupervised method has been known as an adaptable clustering method that can learn while producing the partition of VM objects based on each group similarity as possible and accurate. The proposed unsupervised clustering VM selection based on K-Means or FCM methods can only group the candidate VM based on the number of the clusters reallocated to another physical machine.

Therefore, to choose a specifics VM from a group as decision, we used a rule-based with MAX function. This will improve both unsupervised clustering methods with categorical decision capability.

3.1. K-Means as VM cluster

K-means clustering is a simple clustering method that uses distance measurement to determine the group of the data. This simple approach makes this method capable of separating the data quickly [29]. In this research, K-Means will determine each workload VM's cluster in the overloaded physical machine based on RAM and MIPS as distance properties. The workload dataset that contains the features of each VM will be fed into K-Means. There are several steps to perform the K-Means clustering method, as follow [22]:

- Determine the number of clusters (c). This number represents the number of groups after the partition process is performed. Here we used $c = \{3, 4, 5\}$
- Determine the centroid of each cluster. The number of centroids is the same as the number of clusters. The centroids value is initialized using the previous research formula [4, 8].
- Calculate the distance between each VM's features to each centroid using Euclidean distance (Eq. (1)), then group each VM from the workload dataset based on the centroid closest distance.

$$d(x_j, v_i) = \sqrt{\sum_{j=1}^{n} \sum_{x_j \in V_i} (x_j - v_i)^2} \qquad (1)$$

where $d(x_j, v_i)$ is the distance between data x and centroid v_i in each number of data (j).

- Update the value of the current centroid based on the new group of data using the formula:

$$v_i = \sum_{x_j \in V_i} \frac{x_j}{n_i} \qquad (2)$$

Here, v_i is the updated centroid, and n_i is the number of data in the cluster V_i.

- Redo steps 3 and 4 until the grouping result converges when no data is moving toward each cluster.

The result of K-Means clustering is a group of clusters that contain the nearest VM Data to each cluster centroids. Moreover, from the VM groups of clustering results, the VM that should be migrated from the overloaded physical machine will be determined by the if-then condition in Eq. (6).

3.2. Fuzzy C-means as VM cluster

The implementation of FCM in this research, especially in VM selection, can be described below [28]:

- Determine the parameter required for the clustering process, wherein this research using several workload datasets that contain the attributes of RAM and MIPS.
- Initialized the configuration used by FCM such as:

  a. Cluster Number : {3,4,5}
  b. Error rate : 0.01
  c. Weight : 2
  d. Maximum Iteration : 50
  e. Initial Iteration : 1
  f. Fitness Function : 0

- This study will be used several cluster numbers such as 3, 4, or 5. This variation of cluster numbers is the same number used in K-Means. Then, the weight value in FCM should be more than 1 [30]. Therefore, this research used weight value = 2. The maximum iteration is 50. However, if the fitness function or error rate were satisfied, the clustering process would be stopped.
- Once the configuration is initialized, raised the Partition matrix. The partition matrix uses to identify the VM data related to the clusters according to the highest value. The number of lines will match the number of available VM data in the dataset, and the columns are to construct the cluster numbers.
- Determine the centroid for each cluster using the formula below:

$$v_i = \frac{\sum_{j=1}^{n} \left(\mu_{ij}\right)^m x_j}{\sum_{j=1}^{n} \left(\mu_{ij}\right)^m} \tag{3}$$

where in Eq. (3), v_i is the cluster center. μ_{ij} is the data in the partition matrix located in i data and j cluster. Then, x_j is all data in each cluster located in j. m is the weight used in FCM. The weight value determines the fuzzier level of the cluster. The higher the weight value, the fuzzier the clustering result will be [30].

- Calculate the fitness function using Eq. (4) as follow:

$$P^c{}_{FCM} = \sum_{i=1}^{c} \sum_{j=1}^{n} \left(\left(\mu_{ij}\right)^m d\left(x_j, v_i\right)^2 \right) \tag{4}$$

From Eq. (4), n is the number of data, where $d\left(x_j, v_i\right)$ is the distance between data x_j and the cluster v_i centroid. This research used Euclidean distance. $P^c{}_{FCM}$ is the fitness function value in c iteration.

- The final stage of the FCM process is to update the matrix partition (μ), which was previously used in P_1 . The improvement process was conducted by improving each data membership degrees in each cluster in the matrix partition.

$$\mu_{ij} = \left[\sum_{r=1}^{c} \left(\frac{d(x_j, v_i)}{d(x_j, v_r)} \right)^{\frac{2}{m-1}} \right]^{-1} \tag{5}$$

The equation above (Eq. (5)) is used to improve membership degree, where:

μ_{ij} : The membership degree.

$d(x_j, v_i)$: Euclidean distance between data x_j to cluster center v_i.
m : Weight or hyper-parameter for fuzzier level determination.

- Repeat the process above until the difference between fitness function value i and $i-1$ is smaller than the error rate, or the maximum iteration number is satisfied. The partition matrix will continuedly to be processed until the exceed condition is fulfilled, then the final result of the partition matrix is used.

After the final result of FCM was obtained, the clustering result will be processed using the Rule Base function in Eq. (6) to select the migrated VM from an overloaded host.

3.3. Rule-based MAX function as VM decision

After the final result of K-Means and FCM clustering was obtained, VM data selection was made by selecting the VM with the most extensive index in each cluster.

$$VM = \begin{cases} x_j \in \min(V_i) \leftrightarrow V_i \neq \emptyset \ , \max(x_j \in V_i) \\ x_j \in \min(V_{i+1}) \leftrightarrow V_i = \emptyset, \max(x_j \in V_{i+1}) \leftrightarrow V_{i+1} \neq \emptyset \end{cases} \tag{6}$$

The Eq. (6) is used as the decision maker to select the VM where x_j is the candidate VM instance with index j in an overloaded host that will be moved. The first condition in Eq. (6) describes that it takes the smallest cluster V with index i, and if the cluster V_i is not empty ($V_i \neq \emptyset$), then select VM x with the maximum index j ($max(x_j \in V_i)$) from the cluster (V_i). While in the second condition, if the smallest cluster index V_i is empty, then it takes the VM x_j data with maximum index j ($max(x_j \in V_{i+1})$) from the next smallest cluster index V_{i+1}.

4. Performance Evaluation

4.1. Workload configuration

Workload Dataset provided by PlanetLab [21] is the dataset gathered from real use of CPU VM workload in the data center that consists of 500 servers which collected every five minutes in 24 hours. In this study, we only used data collected on 03-03-2011 (workload A) that contains 1052 CPU VM, 06-03-2011 (workload B) contains 898 CPU VM, and 09-03-2011 (workload C) contains 1061 CPU VM. The distribution of the average VM CPU utilization in each quartile could be seen in Table 1.

Table 1. Distribution of average VM CPU utilization (%).

Workload Dataset	Q1	Q2	Q3	Q4
Workload A	2.76	6.83	14.38	94.59
Workload B	3.08	6.30	11.98	95.98
Workload C	2.84	6.59	12.75	92.65

The quartile 1 (Q1) represents the minimum distribution of the average VM CPU utilization, and quartile 4 (Q4) represents the maximum distribution of the average CPU utilization in 24 hours. This research used original dataset without executing any pre-processing phase to keep the entire range of CPU utilization workload in each VM at each time point.

4.2. VM Instance configuration

There are four different categories of workload specifications in each VM, especially in the heterogeneous VM characteristic. These categories consisted of a Micro CPU with 613 MB of RAM and 500 MIPS, a Small CPU with a 1.7 GB of RAM and 1000 MIPS, and a medium CPU with 1,7 GB of RAM and 2000 MIPS, and a High CPU with 0.87 GB of RAM and 2500 MIPS.

4.3. Data center configuration

The simulation of both methods was conducted using the CloudSim tool to create the data center environment simulation. Each method will be evaluated using three clusters: three clusters, four clusters, and five clusters. The data used to evaluate both methods consist of 800 samples of homogeneous physical hosts with series HP ProLiant ML 110 G5 HP ProLiant ML 110 G5 Xeon 3075. The physical hosts have 2.660 MIPS, 2 PES, 4.096 RAM, 1Gbit/s, 1000 GB Storage.

4.4. Evaluation

Both unsupervised methods will be compared with the robust VM selection method based on the statistical approach used in the cloud data center, such as MMT and RC [20]. Evaluation using Energy Consumption (EC) is proposed to determine the most energy-efficient method implemented in a cloud data center.

5. Result and Discussion

This research has simulated and evaluated the proposed unsupervised VM selection methods using K-Means and Fuzzy C-Means in the CloudSim environment. In this experiment, firstly, we cluster each VM in several workload datasets into different cluster numbers (3, 4, and 5) using K-Means and FCM and several comparison methods (MMT and RC). The total number of workloads VM used in this study is up to 3011 VM collected from 500 servers in 24 hours. After the clustering result was obtained, perform the proposed rule base function to move the VM with the most extensive index in each cluster. After that, we measure the total Energy Consumption (EC) from each workload dataset. Moreover, to measure the significant energy-efficiency level, this study performs a Friedman Test as a non-parametric statistical method.

Energy Consumption (EC) describes the energy used by the cloud data center in kWh (kilowatt-hour). The higher the kWh value, the more extravagant the cloud data center in using energy. The evaluation result of both methods in several cluster numbers can be seen below:

Each method with a different number of the cluster has been tested in three workload datasets, as shown in Figs. 2, 3, and 4. In the workload A dataset (Fig. 2), the highest energy consumption value was presented by MMT with 179.53 kWh of energy consumption. Meanwhile, in the same dataset, the smallest energy consumption was achieved by K-Means with five clusters, where the amount of energy used was 163.02 kWh. Here, the proposed K-Means with 5 clusters can reduce the MMT energy consumption by up to 16.51 kWh. The proposed FCM method in the workload A dataset shows that the method with three clusters performs better than other proposed cluster numbers in the FCM method. FCM with three cluster members achieves 163.83 kWh of energy consumption, which better than the MMT energy consumption of less than 15.7 kWh.

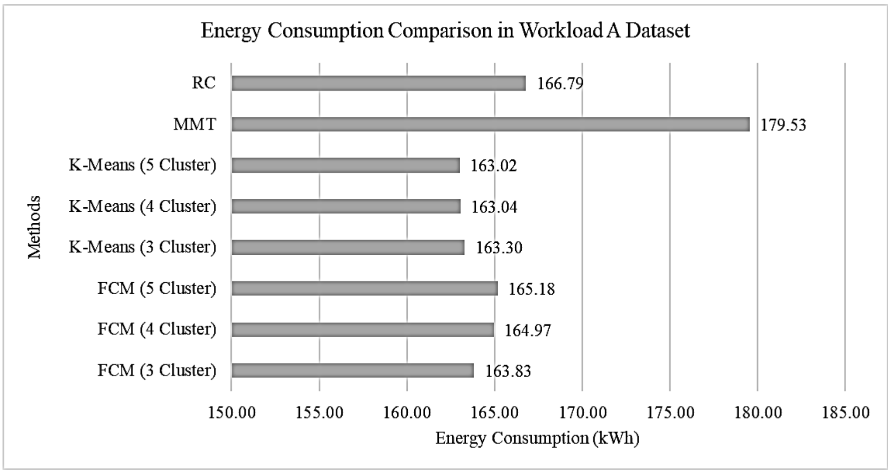

Fig. 2. Comparison of energy consumption in each method in workload A dataset.

Moreover, Fig. 3 describes the comparison of each method's energy consumption in the workload B dataset. This result shows that K-Means with five cluster members produce the lowest energy consumption than other tested methods with 124.44 kWh of energy. Meanwhile, in the FCM method, the lower energy consumption was achieved by FCM with four clusters members with 124.75 kWh of energy. However, the MMT method also presents the highest amount of energy consumption in the workload A dataset. The amount of energy used by MMT is 134.94 kWh. The difference number of K-Means with five clusters and FCM with four clusters to MMT, where MMT is the most extravagant method of all tested methods, is 10.5 kWh and 10.19 kWh, respectively.

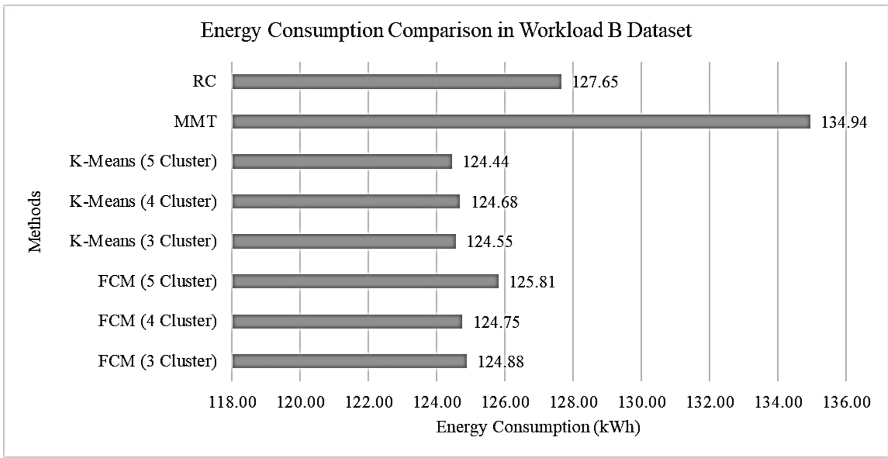

Fig. 3. Comparison of energy consumption in each method in workload B dataset.

Furthermore, the evaluation result using the workload C dataset was described in Fig. 4. The inefficiency energy was achieved by the MMT method with the consumption value of 157.44 kWh. K-Means gained the most efficient energy consumption in the workload C dataset with four cluster members, 144.21 kWh of energy consumption. Then, FCM with four clusters consumes 144.48 kWh of energy in the workload C dataset, where this amount of energy is more efficient than another cluster number in FCM methods. Comparing the energy consumption value of the MMT method to K-Means and FCM with four clusters indicates the difference number equal to 13.23 kWh and 12.96 kWh of energy usage, respectively.

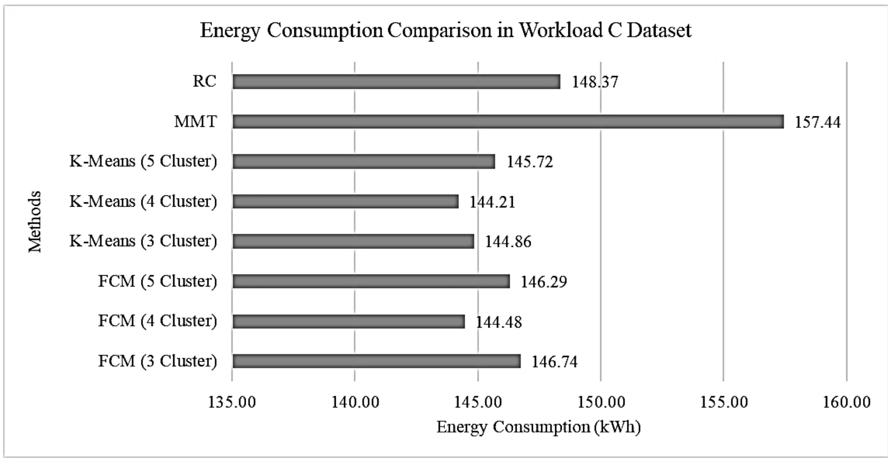

Fig. 4. Comparison of energy consumption in each method in workload C dataset.

The overall result is described in Figs. 2, 3, and 4 show that the unsupervised methods, either K-Means or FCM, provide better energy-efficiency than MMT and RC. This condition occurs since the natures of unsupervised machine learning methods can learn and adjust their weight. This ability gives better adaptation in selecting candidate VM to achieve optimal cluster that is decided by rule-based function. The proposed unsupervised methods can select the best VM that should be migrated from an overloaded host than statistical methods such as MMT and RC.

The significant difference in energy-efficiency results between each method can be determined by a non-parametric statistical evaluation known as the Friedman Test [31]. The null hypothesis H0 used in this research was that there is no significance in energy efficiency among all tested methods. The other possibility of H1 was that there is significant energy-efficiency between all methods.

In conducting the Friedman Test, this research was used all tested methods as treatments. Each treatment has three energy consumption data from three workload datasets, shown in Table 2.

Table 2 shows the Friedman Test's average rank result from three workload datasets that evaluate the VM selection methods. The first rank in the Friedman Test will be defined as the method that provides the highest energy-efficient.

Table 2. Result of Friedman rank.

Method		Energy Consumption (kWh) in each Workload			Average Energy Consumption (kWh)	Sums of Friedman Rank	Rank
		A	B	C			
FCM	**(3 Cluster)**	163.83	124.88	146.74	145.15	15	4
FCM	**(4 Cluster)**	164.97	124.75	144.48	144.73	11	3
FCM	**(5 Cluster)**	165.18	125.81	146.29	145.76	17	5
K-means	**(3 Cluster)**	163.30	124.55	144.86	144.24	8	2
K-means	**(4 Cluster)**	163.04	124.68	144.21	143.98	6	1
K-means	**(5 Cluster)**	163.02	124.44	145.72	144.39	6	1
MMT[6]		179.53	134.94	157.44	157.30	24	7
RC[5]		166.79	127.65	148.37	147.60	21	6

Table 2 shows that K-Means with four and five clusters achieve the first ranks, followed by three clusters in the second position. Besides that, FCM with four, three, and five clusters members could reach third, fourth, and fifth positions in Friedman rank. Moreover, the two last positions were achieved by RC and MMT in the Friedman rank, respectively. The difference number of sums of rank in the Friedman test between each method is significantly huge.

K-Means looks promising that able to show the first rank in the Friedman Test compared to FCM. This condition occurs because of the capability of K-Means to determine better quality and accuracy clusters than FCM [9]. The p-value obtained from the Friedman Test was 0.0105, which is lower than the alpha value $\alpha = 0.05$. This result indicates that the hypothesis H0 was rejected, and H1 was accepted. Thus, the evaluation using the Friedman Test represents that there is significant energy-efficiency among all methods.

6. Conclusion

This research proposed VMs Selection using an unsupervised method in the cloud data center. The evaluation of the proposed method has been conducted using CloudSim. The proposed unsupervised VM selection (K-Means and FCM) gives better energy consumption than the statistical approach (MMT and RC) from the energy-consumption evaluation. The result has shown that all K-Means clusters could reach the lowest energy consumption among all tested methods (FCM, MMT, and RC). The significant improvement of energy-efficiency was proven using the Friedman Test that shown 0.0105 of the p-value, where the first rank of the Friedman Test was achieved by K-Means, the second rank by FCM, the third and last rank was achieved by MMT and RC, respectively.

This research had integrated unsupervised methods with the rule-based MAX function. This function is used to determine the moved VM from each cluster generated by the unsupervised method to improve the energy efficiency of the cloud data center. However, there needs to be further review and evaluations related to the improvement of rule-based methods that can adapt to complex features.

7. Acknowledgements

This research has been funded by “Kementerian Riset dan Teknologi/ Badan Riset dan Inovasi Nasional” with grant id 044/A.38.04/UDN-09/IV/2020 and 014/LL6/PG/SP2H/PL.I/2020 and 8/E1/KPT/2020.

Nomenclatures

c	Cluster number
$d(x_j, v_i)$	Euclidean distance between data x and centroid v_i in each number of data (j).
m	Weight
n	Total number of data
$P^c{}_{FCM}$	The FCM fitness function value in c cluster iteration
V_i	Cluster (i)
v_i	Cluster centroid (i)
x_j	Cluster data (j)

Greek Symbols

μ_{ij}	The data in the partition matrix located in i data and j cluster

Abbreviations

CPS	Constant Position Selection
DVFS	Dynamic Voltage and Frequency Scaling
FCM	Fuzzy C-Means
FQL	Fuzzy Q-Learning
ICT	Information and Communication Technology
MAD	Median Absolute Deviation
MC	Maximum Correlation
MIPS	Million Instructions Per Second
MMT	Minimum Migration Time
RAM	Random Access Memory
RC	Random Choice
SLA	Service Level Agreement
VM	Virtual Machine

References

1. Abbasi, A.A.; Abbasi, A.; Shamshirband, S.; Chronopoulos, A.T.; Persico, V.; and Pescapè, A. (2019). Software-defined cloud computing: a systematic review on latest trends and developments. *IEEE Access*, 7, 93294-93314.
2. Bermejo, B.; Juiz, C.; and Guerrero, C. (2019). Virtualization and consolidation: a systematic review of the past 10 years of research on energy and performance. *Journal of Supercomputing*, 75(2), 808-836.
3. Koomey, J. (2011). Growth in data center electricity use 2005 to 2010, *Analytics Press*, 1-24.
4. Shidik, G.F.; Azhari, A.; and Mustofa, K. (2015). Evaluation of selection policy with various virtual machine instances in dynamic vm consolidation for energy efficient at cloud data centers. *Journal of Networks*, 10(7), 397-406.
5. Beloglazov, A.; Buyya, R.; Lee, Y.C.; and Zomaya, A. (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*, 2(13), 1397-1420.

6. Beloglazov, A.; Abawajy, J.; and Buyya, R. (2012). Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. *Future Generation Computer Systems*, 28(5), 755-768.
7. Masoumzadeh, S.S. (2013). Integrating VM selection policies in dynamic VMs consolidation using fuzzy Q-learning, 332-338.
8. Shidik, G.F.; and Ashari, A. (2014). Efficiency energy consumption in cloud computing based on constant position selection policy in dynamic virtual machine consolidation. *Advanced Science Letters*, 20(10-12), 2119-2124.
9. Panda, S.; Sahu, S.; Jena, P.; and Chattopadhyay, S. (2012). Comparing fuzzy-C means and K-means clustering techniques: A comprehensive study. *Advances in Intelligent and Soft Computing*, 166 AISC(VOL. 1), 451-460.
10. de Carvalho, F. de A.T.; Lechevallier, Y.; and de Melo, F.M. (2012). Partitioning hard clustering algorithms based on multiple dissimilarity matrices. *Pattern Recognition*, 45(1), 447-464.
11. Ferreira, M.R.P.; de Carvalho, F. de A.T.; and Simões, E.C. (2016). Kernel-based hard clustering methods with kernelization of the metric and automatic weighting of the variables. *Pattern Recognition*, 51, 310-321.
12. Bonis, T.; and Oudot, S. (2018). A fuzzy clustering algorithm for the mode-seeking framework. *Pattern Recognition Letters*, 102, 37-43.
13. Tang, Y.; Ren, F.; and Pedrycz, W. (2020). Fuzzy C-Means clustering through SSIM and patch for image segmentation. *Applied Soft Computing*, 87, 105928.
14. Faisal, M.; Zamzami, E.M.; and Sutarman. (2020). Comparative analysis of inter-centroid k-means performance using Euclidean distance, Canberra distance and Manhattan distance. *Journal of Physics: Conference Series*, 2020, 1566(1).
15. Polat, K.; and Senturk, U. (2018). A novel ML approach to prediction of breast cancer: combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier. *2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2018, 1-4.
16. Preeti; Kaur, R.; and Singh, D. (2020). Performance evaluation of clustering techniques in wireless sensor networks. *Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*, 2020, 8(1), 97-103.
17. Angelin, B.; and Geetha, A. (2020). Outlier detection using clustering Techniques-K-means and K-median. *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, (ICICCS), 373-378.
18. Rustam, Z.; and Fijri, A.L. (2020). Breast cancer clustering using modified spherical K-Means. *Journal of Physics: Conference Series*, 1490, 012028.
19. Zhou, K.; and Yang, S. (2020). Effect of cluster size distribution on clustering: a comparative study of k-means and fuzzy c-means clustering. *Pattern Analysis and Applications*, 23(1), 455-466.
20. Ismaeel, S.; Karim, R.; and Miri, A. (2018). Proactive dynamic virtual-machine consolidation for energy conservation in cloud data centres. *Journal of Cloud Computing*, 7(1), 1-28.

21. Park, K.; and Pai, V.S. (2006). CoMon: a mostly-scalable monitoring system for PlanetLab. *ACM SIGOPS Operating Systems Review*, 40(1), 65.
22. Shidik, G.F.; Sulistyowati, N.S.; and Tirta, M.B.W. (2016). Evaluation of cluster K-Means as VM selection in dynamic VM consolidation. *Proceedings - Asia-Pacific Conference on Communications, APCC 2016*, 124-128.
23. Beloglazov, A. (2013). *Energy-efficient management of virtual machines in data centers for cloud computing*. PhD thesis. Department of Computing and Information Systems, The University of Melbourne.
24. Ashraf, A.; and Porres, I. (2018). Multi-objective dynamic virtual machine consolidation in the cloud using ant colony system. *International Journal of Parallel, Emergent and Distributed Systems*, 33(1), 103-120.
25. Karthikeyan, K.; Sunder, R.; Shankar, K.; Lakshmanaprabu, S.K.; Vijayakumar, V.; Elhoseny, M.; and Manogaran, G. (2018). Energy consumption analysis of Virtual Machine migration in cloud using hybrid swarm optimization (ABC–BA). *Journal of Supercomputing*, 76, 3374-3390.
26. Dinesh Reddy, V.; Gangadharan, G.R.; and Rao, G.S.V.R.K. (2019). Energy-aware virtual machine allocation and selection in cloud data centers. *Soft Computing*, 23(6), 1917-1932.
27. Mustofa, K.; Azhari, A.; and Shidik, G.F. (2018). K-mMA VM selection in dynamic VM consolidation for improving energy efficiency at cloud data centre. *International Journal of Communication Networks and Distributed Systems*, 21(2), 202.
28. Shidik, G.; Noersasongko, E.; Nugraha, A.; and Setiono, O. (2019). Migrating virtual machine in heterogen and homogen cloud data center with efficient electricity consumption using fuzzy C-meams. *International Journal of Intelligent Engineering and Systems*, 12(6), 199-206.
29. Singh, M.; and Bansal, M. (2014). Survey on clustering and optimization techniques to develop hybrid clustering technique. *International Journal of Computer Engineering and Applications*, VII(I), 93-101.
30. Zhang, J.; and Shen, L. (2014). An improved fuzzy C-means clustering algorithm based on shadowed sets and PSO. *Computational Intelligence and Neuroscience*, 2014, 1-10.
31. Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200), 675-701.