



An Efficient Technique for VM Database Segmentation over Cloud Environment

SUNILMANOI¹, S CLINGAREDDY²

¹Research Scholar, School of Computer Science and Engineering, REVA University, Bangalore India

² Professor, Computer Science and Engineering, SRI Venkateswara College of Engineering Technology Bengaluru, India

* Corresponding author's Email: sunil.manoli@reva.edu.in

Abstract:

Cloud computing models use virtual machine (VM) clusters for protecting resources from failure with backup capability. Cloud user tasks are scheduled by selecting suitable resources for executing the task in the VM cluster. Existing VM clustering processes suffer from issues like preconfiguration, downtime, complex backup process, and disaster management. VM infrastructure provides the high availability resources with dynamic and on-demand configuration. The proposed methodology supports VM clustering process to place and allocate VM on the basis of requesting task size with bandwidth level for enhancing the efficiency and availability. The proposed clustering process is classified as preclustering and postclustering on the basis of migration. Task and bandwidth classification process classifies tasks with adequate bandwidth for execution in a VM cluster. The mapping of bandwidth to VM is done based on the availability of the VM in the cluster. The VM clustering process uses different performance parameters like lifetime of VM, utilization of VM, bucket size, and task execution time. The main objective of the proposed VM clustering is that it maps the task with suitable VM with bandwidth for achieving high availability and reliability. It reduces task execution and allocated time when compared to existing algorithms.

1. Introduction

both general purpose computers and data networks A method of service-oriented architecture known as "cloud computing" leverages virtualized resources to carry out computations. A number of resources are available in the cloud as a way of service. The different types of cloud services include SaaS, IaaS, and PaaS. To satisfy customer needs, many deployment models are used for the services. They fall under the categories of private, public, community, and hybrid clouds. While services are provided as a public cloud outside of the organization, private cloud

resources are shared within the organization. A sort of cloud called community cloud allows service providers in the same category to share resources. A hybrid cloud, which combines two or more deployment types, offers the customer service. Cloud services are modeled by mapping the virtualization

layer to the appropriate VM. VMs are selected from the VM list and then mapped to the respective request generated by the user of the cloud service.

The cloud consists of a heterogeneous host in a data center that maintains a mobile resource-based access environment.

VM accesses that leads to a performance problem seen in the areas of the battery life and energy consumption. The entire performance factor in green computing is made use of to overcome this problem [1]. Mobile cloud computing (MCC) is a mobile resource sharing service that allows mobile devices to have access to the appropriate cloud service. It has faced challenges in terms of scalability, such as computational storage services and other different services [2].

Mobile computing over cloud has the ability to target parameters such as traffic, quality, and customer demand. Traditional static cloud and dynamic cloud are compared to analyze the workload. The static cloud allows users to access infrastructure services with a specific configuration, while the dynamic cloud provides an agile response method to update the resource configuration.

Dynamic cloud has a variety of wireless nodes with device-to-device connectivity for the achievement of a better utilization of the channel and traffic [3]. VM contention: IaaS has created problems in the area of performance.

This problem is overcome through the implementation of the data center as various ranges such as single server with virtualization, single mega data center, and multiple geo-distributed data centers [4]. A researcher engaged in cloud suffers from issues that include energy consumption in data centers. The data center is a key element in the cloud that handles all kinds of

resources needed in the computing environment. In the cloud, there are 2 types of approaches that are related to hardware and software. These approaches require reduced power consumption in cloud resources without any situation of service unavailability [5]. Cloud resources are multiplexed over VM servers designed to host cloud services in large data centers.

VM migration is the process of migrating from one location to another leading to the performance problem arising because of inference and the cost of the operation. iAware imposes the multi-source supply

demand model for minimizing the inference during the migration [6, 7]. Intelligent transportation systems (ITS) are used in the vehicle cloud computing (VCC) architecture, which consists of two-layered components such as the central cloud server and the remote system control (RSC). RSC is a remote administration manager for monitoring and managing distributed system elements such as network communication lines. RSC uses two local server and road side unit (RSU) components. If the vehicle travels from one position to another, the VM of one RSU is moved to another RSU. It requires continuous service response over the automated vehicle control using VCC [5]. There are two main disputes seen in the intelligent transportation systems (ITS), namely, efficiency in traffic and energy, quality, and productivity. The data collected from various sensors are overcome by using the parallelized fusion technique.

This technique follows the Dempster–Shafer theory with four components namely, sensor input, bootstrapping, hierarchical fusion, and state output [6, 8]. Cyber-ITS is a system that has data division, scheduling, and efficient support through the use of a generic methodological framework. There are two types of functions carried out by the system, namely, data-centric and operation-centric transformations. This model uses high-performance Computer design with region-based decoupling capability [9, 10].

In this method, a digital map of the global positioning system data is processed in a parallel manner using quad tree-based domain decomposition technique. These data are divided into different subdomains with quad structure [11]. Multi-CPU VM scheduling and virtual CPUs (VCPUs) scheduling have been carried out due to the availability of various virtualization techniques in the cloud computing domain [10]. The existing problems have been analyzed on the basis of performance parameters, which improve the efficiency in the cloud service deployment. The main objective of the proposed technique is to identify the suitable VM on the basis of bandwidth and requesting task of allocating any issues in the performance of cloud to the task.

The VMs are configured in an isolated fashion, which suffers from repetitive booting of the respective volumes with a limited period of delivery. This problem is solved by a cluster management approach based on a Docker container with a diverse configuration [12]. The traditional placement of the Docker container and the VM is carried out separately, so that it is implemented using the container VM-PM model [13]. The Internet of Things plays a key role in the processing of real-time data from hardware devices that generate large quantities of data. These data are stored in a large data center with Big data analysis methods. If the data are huge, a huge number of servers would be used to store the received data. It

faces an expensive problem that is solved by using a cabinet-based tool called ProCon [14].

Virtualization technology offers the benefits of the physical server operating on several computers with different resources. Virtual storage eliminates read and write delays with high I/O efficiency. Virtual disks are connected to VMs for processing and storing the user data via synchronization [15]. Amazon cloud provider provides various kinds of services to the end user in reliable and secure computing capacity. AWS offers the Elastic Compute Cloud (EC2) with different versions of VM and resources. The proposed method used EC2 instances as a reference for further analysis.

2. Proposed Methodology

Existing VM management techniques configure a VM to cloud workloads based on the energy parameters, but they suffer from a resource wastage problem in the data center. The proposed model groups the tasks based on the VM types with bandwidth parameters using classification methods. The VM types are customized on the basis of resource availability in the cloud. The objective of the proposed model is to map a task to the correct VM by considering various processes such as resources mapping and classification [16]. The cloud requesting tasks are classified based on the metrics such as total number of queues, request count, API response count, dispatch rate of the queue, maximum size of the task, actual and scheduled execution time, delay, and task retention time. These metrics are collected using Cloud Watch monitoring service in AWS cloud. These metrics are exported and used for analyzing the proposed model.

Figure 4.1 shows the architecture of the proposed system. The customer makes a request for the resource from the cloud based on their current requirement. The request is considered as a task executed in a VM. There are various tasks generated by the customer identified based on various perspectives. The identified task is classified on the basis of reliability parameter for achieving a high performance. Bandwidths are selected for the classified tasks so that a suitable VM is allocated to the requesting task. There are various bandwidths available in the requested task, which are identified and classified to enable mapping the suitable VM for service delivery. The bandwidth-VM mapping section selects the VM from the VM clusters, which, in turn, selection is done by the VM selector. The VMs are clustered on the basis of the task and bandwidth meant for providing the services without any interruption or delay. They are clustered in the respective VMs based on the type of task requested, and the hypervisor eventually provides the link between the VMs and the physical system.

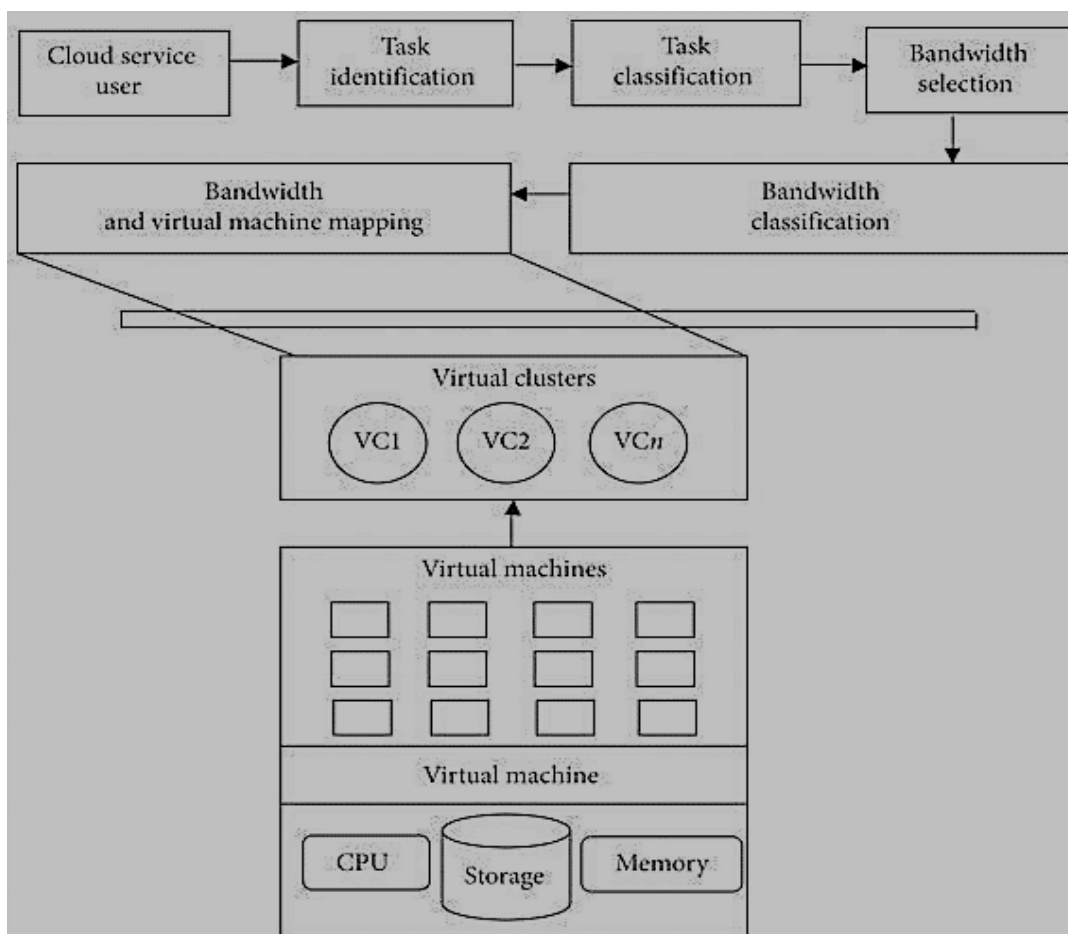


Figure1:Architectureoftheproposedsystem.

Figure 2 is the sequence diagram of the proposed system. It denotes how the incidents are actually related to each other. The activity diagram shows how the process starts and terminates, the various state changes, and activities that take place between these state changes. The first phase is the login module, where the user is authenticated to access the system. The username and password are provided. These further provide access to the cloud homepage, which consists of all the main functionalities. This module allows only the authorized user to log in to the system. It provides authentication and access only to the authorized user and allows the user to select any of the options that include creation of a virtual machine, viewing the existing machines, making task-based separation, and viewing the usage report. The first option in the module is helpful for the user in creation of the module by just entering the values for the new VM. Instead of typing commands in the terminal to create a virtual machine, this module helps us create a VM by simply entering the values.

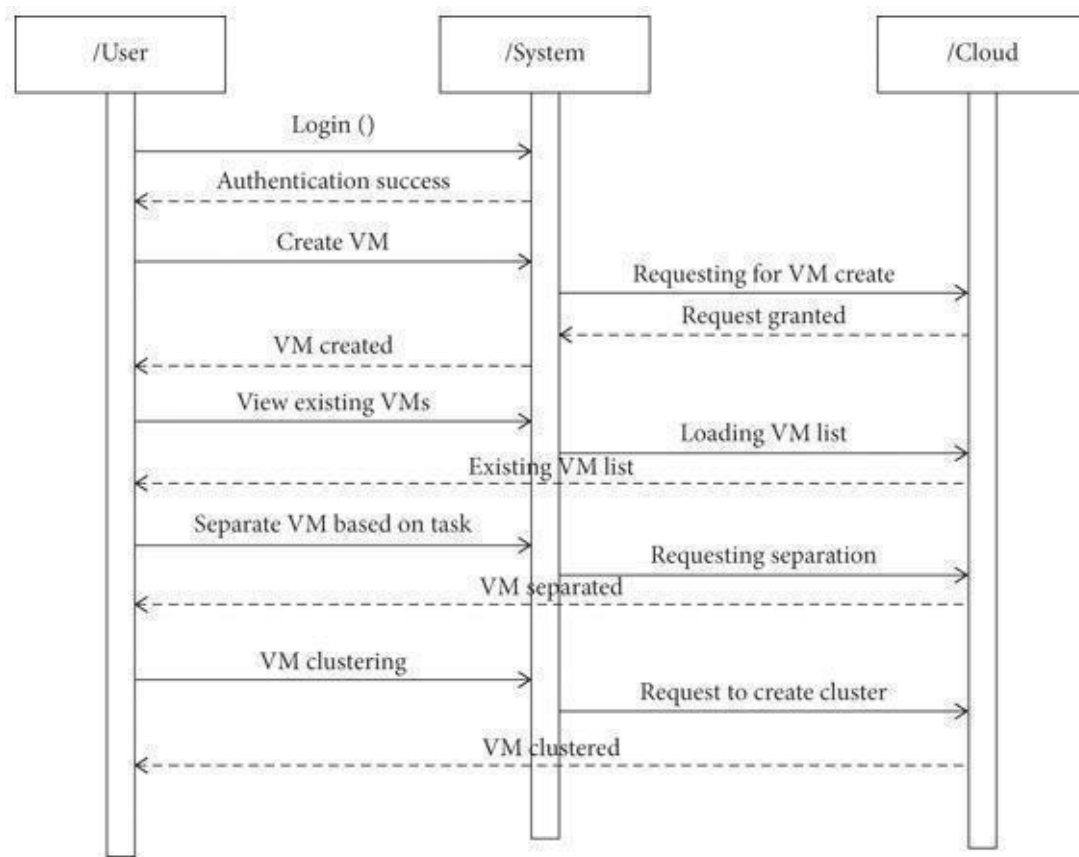


Figure2: Phases of the proposed VM clustering process.

The second option in the module is for viewing the existing VMs, thereby enabling the user to view the type of VM created and used. Each VM has different specifications. This module can view the existing VMs with their specification, the operating system of the VM. The third option in the module is task-based separation. This phase helps separation of the task from the cloud user with the help of this option. It separates the task in terms of CPU, memory, and IO. The fourth option in the module is VM clustering. Groups of VMs with similar features are clustered with the help of this option. The final option is the report, where the user can view a detailed usage of the VMs. In this module, the user has the ability to create new VMs for use based on requirements. The major use of this module is that, unlike in the normal VM creation where the user needs to go to the terminal and type commands, the user just needs to type values for the VM creation. A name for the VM and the various parameters such as the number of CPU, vCPU, the disk name, NIC, and SSH are entered, where all these parameters are required for the creation of the VM template.

The template, which is created, is instantiated by providing the disk name and the VM name, thereby creating the VM. This module is used for helping the user in making separation on the

VM based on the type of task in each request. There are three major classifications on the task that include memory-based, CPU-based and IO-based classification. The CPU-based classification

is meant for the user who requires high processing speed during the processing of the input file. In this, the input file is only processed, but not stored. The memory-

based classification is for the user to have better memory space. In this, the input file is only stored in the memory and not processed. The IO-based classification is used for the user to have a responsive VM. In this option, the output is only generated for the input file and not stored in the memory.

The mixed option is used for the user in the classification of a task with more than one type. Selection of this type helps the user

in the selection of either of the two types of classification methods for each task. This module is used for allowing the user to view the existing VMs that are created in the system. This view option is in a tabular view, where all the existing VMs are listed along with the specifications of each VM. The module is highly useful for getting knowledge of all the VMs that have already been created in the system. All the active VMs can be viewed with the help of this module.

The basic uses of this module provide a view of the existing VMs and differentiate active VMs. The module displays the information of the existing virtual machines such as their user, ID, group, and name.

2.1 Analysis of Various Performance Metrics

2.1.1 VM Parameter-Level Analysis

There are different features considered while clustering the VM for a good maintenance of reliability in the cloud. The capabilities available in the migration process need to keep the copy of the VM at the original source end. Two types of copying process happen in the VM, namely, *recopying* and *postcopying* process. The cluster process uses these techniques for achieving better availability. The CPU is halted during the migration processing the source machine, as well as the destination

machine. Delta-based compression of VM has more number of dependent VMs with respective references for future VM consolidation at the target machine. Data-level compression is used for getting the reduction in the contents related to the VM at the source machine. The workload is classified as synthesis-based and idle-based workload. Synthesis-based workload is a pre-allocated load assigned before the VM migration. Idle-based workload is assigned to the task on the basis of the demand in nature. VM size depends upon various factors, namely, vCPU number, memory size in GB, the bandwidth of the memory in GB/s, the frequency of the CPU in GHz, single and all core frequencies in GHz, performance of remote memory access, temporary storage in GB, number of data disks, and number of Ethernet NICs. Initially, the target machine memory is considered as dirty pages. Later, it is replaced with the respective contents after migration. Resource availability of the target machine is also addressed during the migration of the VM. Table 1 shows the VM features in multiple perspectives.

2.1.2 User Task Classification-Level Analysis

Cloud user tasks are classified on the basis of factors that include the name of the user base (UB), regions and the number of requests per single user, requested task size in bytes, duration in terms of peak hours in GMT, and the average number of peak users offline and online. Table 1 shows the classification of tasks with user parameters of different sizes. These tasks are mapped with the VM parameters for achieving high reliability. Cloud regions maintain the resources for users with the corresponding user base. The requesting size and related task size are analyzed in order to find an average number of users in both peak and non-peak hours of the particular region. The comparison of task classification is shown in Table 2

UB2	1	60	90	4	10	1000	100
UB3	2	40	75	5	8	1000	100

UB4	3	55	80	7	11	1000	100
UB5	4	30	85	8	9	1000	100
UB6	5	45	100	6	12	1000	100

Table2:Thecomparisonoftaskclassification

Cloud region collects the user requests from the user base (UB) and identifies the request size. The tasks are allocated to the respective resources in the cloud region based on the peak.

Heterogeneous VMs are available for handling the user-requesting task with different categories. Virtual machine managers schedule the task to respective VM. A large number of requesting tasks to be handled have arrived in the cloud that are performed by a large number of VMs. It is proposed that the number of virtual machines be increased on the basis of number of tasks requested and non-peak hours through auto scaling technique. It handles the maximum number of VM on the basis of demand. Figure 4.3 shows the total number of tasks on the VMs over the cloud region.

Table4.3: User task classification and comparison.

Name of the user base	Regions	Requests per user per	Task size per request (bytes)	Peak hours starts (GMT)	Peak hours end (GMT)	Avg peak users	Avg off-peak users
UB1	0	60	100	3	9	1000	100

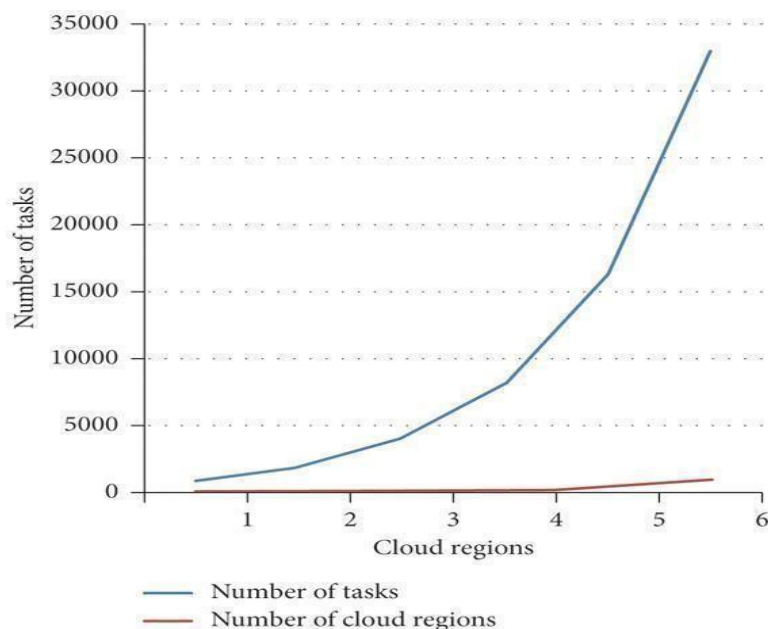


Figure3: Comparison of cloud regions and user tasks.

2.2 Virtual Machine Analysis of Data Centers

Virtual machines are created with different parameters such as the name of the data centers, regions of the data center, the architecture of the VM, the platform of the VM, type of VMM, resource cost, and physical hardware units. Data centers perform various operations, which include migration process in either the same data center with migration or different data centers with migration. Linux with X86 architecture is used for deciding the required number of physical machines at the data centers in different regions. Describes the VM attributes for task allocation and execution. Cloud customer needs higher bandwidth because of the lack in the current bandwidth level. Cloud providers provide sufficient bandwidth in order to retain the customers. Table 3 provides the bandwidth level of various domain users for effective utilization.

Table 3: Comparison of bandwidth utilization.

Domain of service providers	Current bandwidth utilization (%)	Expected bandwidth level (%)

Banking and insurance	61	19
Telecommunications	73	10
IT services	39	16
Cloud services	62	4
Education services	36	4
Government services	41	10

2.3 Delay and Bandwidth Matrix Analysis

Cloud services are deployed in various data centers as regions. Delays seen between the regions are compared for efficiency. The main objective of this delay analysis is to identify the fast response with minimum response rate. The bandwidth matrix provides an efficient route between the requests in the shortest path with maximum availability. Delay in the network during the transfer of jobs across various regions is shown in Table 6. The delay between the same region is also minimum, whereas the delay of the different regions is maximum with respect to the distance between the regions. The same region transfer rate is fast when compared to different regions, so it depends on the bandwidth level and delay, which is shown in Table 7. Figure 4 indicates the response rate of various cloud regions in ms (Milliseconds). Response time of various user bases is measured in three levels, namely, average, minimum, and maximum. The Cloud Analyst model is used for the proposed analysis with delay and bandwidth allocation of the task execution. The custom bandwidth and delay matrix are specified on the internet characteristics option with various regions.

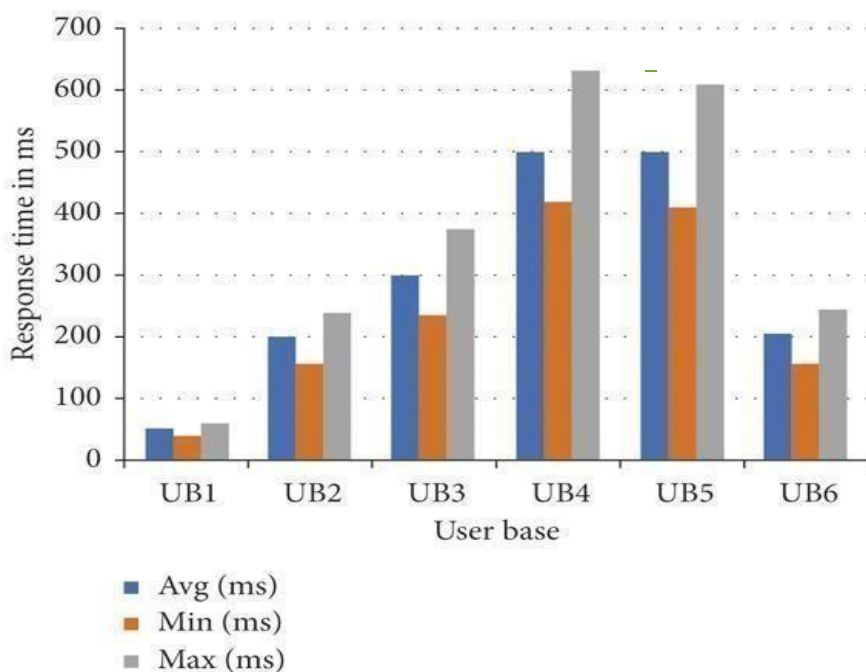


Figure4: Comparison of the response time of the cloud region.

2.4 Analysis of VM Clustering

VM clustering is the process of grouping various VMs using virtual networks for achieving high availability of cloud resources. This clustering process is done in a source machine, as well as a target machine. There are two important concepts considered for good accuracy, namely, pre-clustering and post-clustering. A preclustering process occurs at the source end, whereas post-clustering occurs at the target machine end. Preclustering interconnects the VM along with the state of the processor, data, and VM-related parameters. These VMs then migrate to the target machine. The postclustering process collects the VMs in order and finds the relationships, forming a cluster by interconnecting the received VMs. This process resumes all the states of the VM and related elements into the original state. Table 8 shows the various VM clusters with the required parameters. There are two types of VMs grouped, namely, active and inactive. Utilization of the CPU plays a vital role during the clustering process. VM lifetime is considered in allocating better performance during the clustering process.

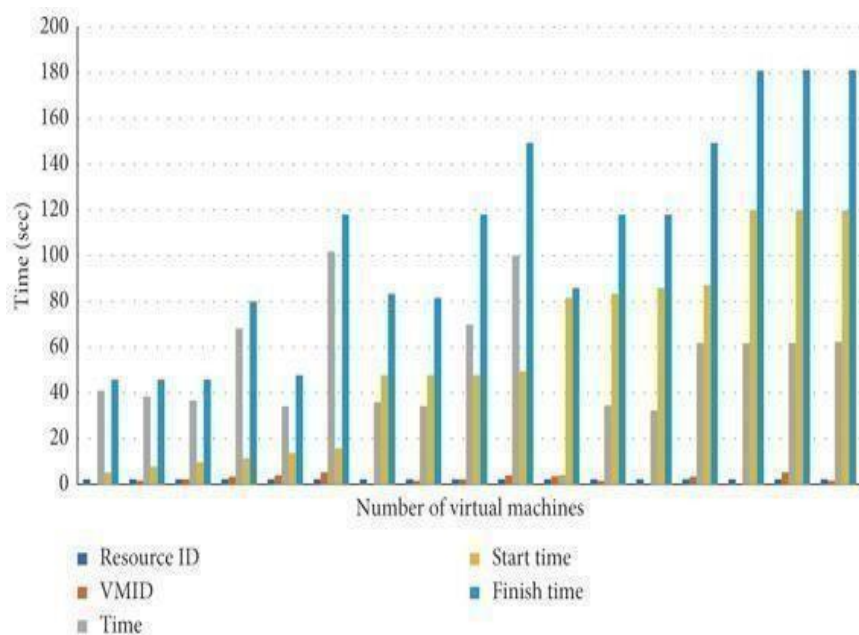


Figure5:TimecomparisonofVMcluster2

2.5 Proposed VM Clustering

The clustered virtual machines (VMs) are created from a physical machine (PM). The mapping of VM and PM is performed by the hypervisor. The objective of clustering is to execute the requesting task, which is categorized with different parameters. The bandwidth is classified, and the corresponding VMs are mapped. The VMs are clustered in a way of similar categories of VM grouped together. The allocation of VMs is extremely simple and efficient for the execution of a task. There are large numbers of clustered VMs that function as dynamic behavior clustering. The completion time of the VM is analyzed, and clusters of VMs are shown in Figures 5-7.

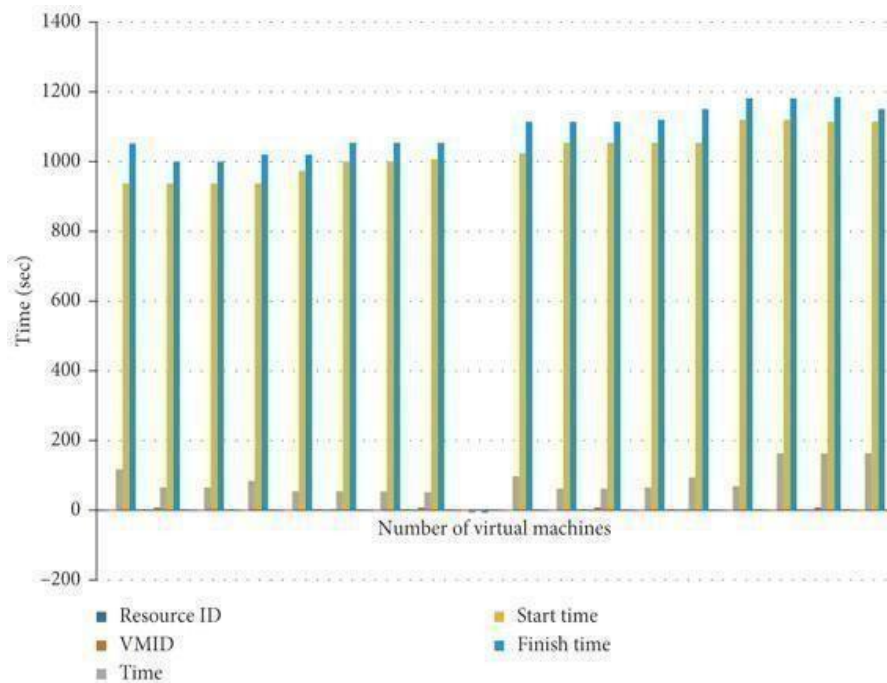


Figure6:TimecomparisonofVM cluster3

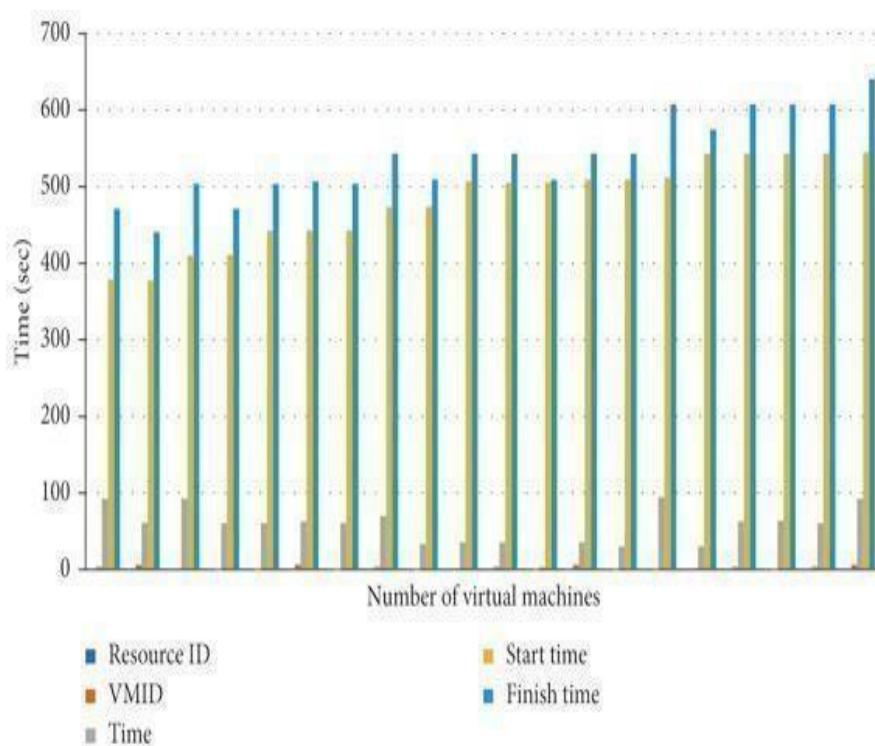


Figure7: Time comparison of VM cluster 4.

5. Conclusion and Future Scope

The focus of the proposed system is on the clustering of VM based on various performance parameters like the type of the requesting task and bandwidth. The requesting tasks are reclassified into CPU-based, storage-based, and IO-based mixed types. To start with, the requesting task is validated, and then, the tasks are allocated to the task classification process. The task classification process categorizes the task into different types depending upon the properties, which exist in the task. These tasks are clustered using the clustering algorithm, in which the categories are grouped together. There are two types of clustering process being carried out, namely, pre-clustering and post-clustering. The bandwidth is also clustered based on the task in the task cluster. There are two types of clusters maintained in the proposed technique, namely, the task cluster and the bandwidth cluster with the same features. The VM is classified and mapped to the suitable requesting task for execution. The proposed technique's main objective is to map the requesting task to a suitable VM, in which the latency in the service handling is minimized with high efficiency.

REFERENCES

- [1] F. Xu, F. Liu, H. Jin, and V. V. Athanasios, "Performance parameters like the type of the requesting task and bandwidth. The requesting tasks are classified into CPU-based, storage-based, and IO-based mixed types. To start with, the 'Managing performance overhead of virtual machines in cloud computing: a survey, state of the art, and future directions,'" *Proceedings of the IEEE*, vol. 102, pp. 11–31, 2013.
- [2] H. Ghiasi and M. Ghobaei-Arani, "Smart virtual machine placement using learning automata to reduce power consumption in cloud data centers," *Smart Computing Review*, vol. 5, no. 6, pp. 532–562, 2015.
- [3] Y. Xia, X. Li, and Z. Shan, "Parallelized fusion on m-transportation data: a case study in CyberITS," *International Journal of Intelligent Systems*, vol. 28, no. 6, pp. 540–564, 2013.

- [4] Y. Xia, T. Zhang, and S. Wang, "A generic methodological framework for cyber-ITS: using cyber-infrastructure in ITS data analysis cases," *Fundamenta Informaticae*, vol. 133, no. 1, pp. 35–53, 2014.
- [5] K. T. Raghavendra, "Virtual CPU scheduling techniques for kernel based virtual machine (KVM)," in *Proc. of IEEE International Conference on Cloud Computing in Emerging Markets*, pp. 1–6, IEEE, Bangalore, India, October 2013.
- [6] Y. Xia, Y. Liu, Z. Ye, W. Wu, and M. Zhu, "Quadtree-based domain decomposition for parallel map-matching on GPS data," in *Proceedings of 15th IEEE Conference on Intelligent Transportation Systems Conference*, pp. 808–813, IEEE, Anchorage, AK, USA, September 2012.
- [7] Y. Mao, J. Oak, A. Pompili, D. Beer, T. Han, and P. Hu, "Draps: dynamic and resource-aware placements scheme for docker containers in a heterogeneous cluster," in *Proceedings of the 2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)*, IEEE, San Diego, CA, USA, December 2017.
- [8] R. Zhang, A. Zhong, B. Dong, and F. Tian, "Container-VM-PM architecture: a novel architecture for docker container placement," in *Proceedings of the International Conference on Cloud Computing*, pp. 128–140, Springer, Seattle, WA, USA, June 2018.
- [9] Y. Fu, S. Zhang, J. Terrero, Y. Mao, G. Liu, S. Li et al., "Progress-based container scheduling for short-lived applications in a kubernetes cluster," in *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*, pp. 278–287, IEEE, Los Angeles, CA, USA, December 2019.
- [10] M. Lee and Y. I. Eom, "Transaction-aware data cluster allocation scheme for qcow2-based virtual disks," in *Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 1–6, IEEE, Busan, Korea, February 2020.
- [11] S. F. Piraghaj, R. N. Calheiros, J. Chan, A. V. Dastjerdi, and R. Buyya, "Virtual machine customization and task mapping architecture for efficient allocation of cloud data center resources," *The Computer Journal*, vol. 59, no. 2, pp. 208–224, 2016.
- [12] Q. Liu, C. Weng, M. Li, and Y. Luo, "An in-VM measuring framework for increasing virtual machine security in clouds," *IEEE Security & Privacy*, vol. 8, no. 6, pp. 56–62, 2010.
- [13] L. Yu and C. Weng, M. Li and Y. Luo, "SNP disk: An efficient para-virtualization snapshot mechanism for virtual disks in private clouds," *IEEE*

- Network*, vol.25,no.4,pp.20–26,2011.
- [14] P.Graubner,M.Schmidt,andB.Freisleben,“Energy-efficient virtual machine consolidation,”*ITProfessional*,vol.15,no.2,pp.28–34,2013.
- [15] M.DuandF.Li,“ATOM:efficienttracking,monitoring, and orchestration of cloud resources,”*IEEETransactionson ParallelandDistributedSystems*,vol.28,no.8,2017
- [16] L. Cui, Z. Hao, and X. Yun, “Piccolo: a fast andefficientrollbackssystemforvirtualmachineclusters,”*IEEETransactionsonParallelandDistributedSystems*,vol.28,no.8,2017.
- [17] X. Bu, J. Rao, and C.-Z. Xu,“Coordinatedself-configurationofvirtualmachinesandappliancesusingamodel-freelearningapproach,”*IEEETransactionsonParallel and Distributed Systems*, vol. 24, no. 4,pp.681–690, 2013.J.Tang,P.T.We,Q.S.Q.Tony,and L.Ben,“systemcostminimizationincloudRANwithlimitedfronthaulcapacity,”*IEEETransactions on Wireless Communications*, vol.16,no.5,2017.
- [18] W. Zhang, Y. Chen, X. Gao, Q. Z. ZhichaoMo,andZ.Lu,“Cluster-Awarevirtualmachinecollaborative migration in mediacloud,”*IEEETransactionsonParallelandDistributedSystems*,vol.28,no.10,2017.
- [19] C. Saravanakumar and C. Arun, “Efficient idlevirtual machine management for heterogeneouscloud using common deployment model,”*KSIITransactionsonInternetandInformation Systems*,vol.10,no.4,pp.1501–1518,2016.\
- [20] J. Vallone, R. Birke, and L. Y. Chen, “Makingneighborsquiet:anapproachtodetectvirtualresourcecontention,”*IEEETransactionsonServices Computing*, vol. 13, no. 5, pp. 843–856,2020.
- [21] S.T.Teixeira,R.Calheiros,andD.Gomes,“Cloud reports: an extensiblesimulationtool forenergy-aware cloud computingenvironments,”in*Cloud Computing.ComputerCommunications and Networks*, Z. Mahmood, Ed.,Springer, New York,NY,USA,2014.