ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 2 FEB 2023

Study of RM Techniques and Performance Evaluation Parameters for Resource Management in Cloud Computing Environment

[1] Om Prakash *, [2] Dr. Muzaffar Azim *, [3] Prof. S.M.K. Quadri

[1] Research Scholar, FTK-Centre for IT, Jamia Millia Islamia,New Delhi, India
[2] System Analyst, FTK-Centre for IT, Jamia Millia Islamia, New Delhi, India
[3] Hony. Director, FTK-Centre for IT, Jamia Millia Islamia, New Delhi, India
Corresponding Authors Email: [1] * opkushwaha2013@gmail.com, [2] * mazim@jmi.ac.in, [3] quadrismk@jmi.ac.in

Abstract*— The cohesive development of Cloud computing attracts organizations and individuals to change their IT infrastructures and environments. According to the varying demand and its development of using Cloud computing, Cloud providers continuously use the upgraded resources of Cloud infrastructure to fit the varying demands. Cloud computing has developed as the main model for delivering and hosting computing resources as services over the Internet. It renders a delivery model for computing resources at platform, infrastructure and software levels. So, Resource Management techniques are proposed in cloud for workload demanding and computing applications that are having different optimization parameters. This survey represents a broad study of Resource Management techniques and refined their huge classification based on the several features. It explores the platforms and evaluation parameters which are utilize to access the Research Management techniques.

***Index Terms*— Cloud Computing, Resource Management, RM Techniques, Performance Evaluation Parameters.**

I. INTRODUCTION

The several varieties of associated computer systems that exist of collaborative and unified computing resource are known as the Cloud. The expansion of cloud computing has enabled to simulate the immediate arrangement of associated data centres that are geographically dissipated for offering steady and higher grade services in the recent years. Cloud computing has pivoted into a decisive prototype to provide computational capabilities on a "pay-per-utilize" presumption in these days. Cloud computing brings abidance and change in the IT businesses. With its evolving application and advancement, cloud computing provides fabulous services, as well as address various difficulties in the growth of traditional IT [3].

The standard service models of cloud computing services are as follows: “Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS)” and Function as a Service (FaaS). These models are mostly categorized based to their offered services. Infrastructure as a Service (IaaS) is a category of Cloud Computing where computing infrastructure is used by customers on a "pay-as-you-go" basis. Often IaaS is considered outsourcing. IaaS is scalable to meet the needs of customers. With IaaS, customers can have all the same computing services as very large firms. In IaaS, consumers hire computing resources (such as storage, compute and network) from cloud service providers. The underlying cloud infrastructure is managed and deployed by the cloud service provider.

Platform as a Service (PaaS) is a category of Cloud Computing that delivers development environments and tools for developing applications. PaaS delivers a pay-as-you-go suite of tools to firms of any size. The development tools can be the top line products that are very expensive.

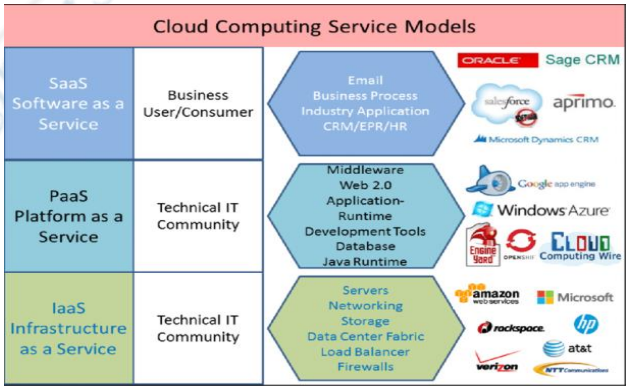

Fig 1. Cloud Computing Service Model

In PaaS, a cloud service typically includes computing, network resources and storage along with an OS, a database, a software development framework, middleware, and tools to deploy, test, develop and manage applications.

Software as a Service (SaaS) is a category to of Cloud Computing where a 3rd party hosts software products, and makes them available to users over the internet. SaaS offers a pay-by-user, and pay-by-month options. These options are much more efficient than the perpetual license model offered by many software vendors. . In SaaS, the entire cloud

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 2 FEB 2023

infrastructure and applications are owned and managed by the cloud infrastructure.

Function as a Service (FaaS) is a very specialized category of Cloud Computing. FaaS delivers real time functionality to firms, when the functionality is needed. With FaaS, the customer does not pay for idle time. During idle time the processing stops. The processing starts again when the user requests the functionality. Only paying for "up time" is a major difference between FaaS, and IaaS. With IaaS, customer pay for the computing infrastructure with it is active, or inactive.

There are two types of resources used in cloud. Physical resources consist of computer, disk, database, network, scientific instruments etc whereas logical resources contain execution, monitoring, communicate applications etc.

The resources of cloud are equipped in the form of virtual machines (VMs) for the customers in IaaS service model. First of all, a Virtual Machine is setup on a physical machine/certain host according to the needed resources on the cloud service provider's data centre. The virtual machines should be capable to execute the tasks of customers or occurring of SLA violation. The tasks of customers may alter in their demand during execution for the cloud resources (increase or decrease). This may trail to overloading/unloading the physical machine that executes the VM of customer. The cloud service providers may opt to migrate or move a VM from one host to another host in order to save in power consumption (in case of host unloading) or ignore SLA violation (in case of host overloading). The cloud service provider must be capable to know when a host is unloaded or overloaded, which and where VMs to migrate them in these scenarios.

Resource integration and management are one of the primary prospects of virtualization technologies engaged in Cloud environments. It is requiring less physical resources than ever before because it permits to a number of standalone physical machines to be consolidated in a virtualized environment within a cluster environment using hypervisors. It often inadequate while improving the situation. Thousands of physical machines and megawatts of power are required for the large Cloud deployments.

The organizations of rest of the paper are as follows. A detail summary of existing surveys is presented in Section 2. The classification of Resource Management techniques are illustrated in Section 3. The major performance evaluation parameters are focused in Section 4. The conclusion presents in Section 5 and Section 6 provides the References. This survey presents the locution of Resource Management techniques, their precise working, evaluation parameters, and exceptive discussions. The representative samples of the most significant work are provided for selected papers. A systematic review strategy and composite methods were used for the selection of paper.

II. LITERATURELP REVIEW

In the recent few years, great observation has been received from the researchers for Research Management techniques for cloud computing environment. Various Research Management techniques are developed which focuses on research challenges which can obstruct smooth planning and upkeep of resources of cloud. Nevertheless, there is very limited review or survey articles are present to assist new and present researchers to understand the functioning and primary concepts of the present methods. In this segment, we talk about and focus on different studies that someway specify to provide brainstorm on Research Management in cloud computing environments.

Saad Mustafa, Atta ur Rehman Khan,Amir Hayat, Babar Nazir, Sajjad A. Madani [1]," provided a detailed survey on Research Management in cloud computing that presented an extensive review of Resource Management techniques and extend their comprehensive aspects depends on the different functions and features. It illustrates the platforms and the various evaluation parameters that are used to figure out the RM techniques. Furthermore, it shows research challenges and design goals that may be recognised during introducing peculiar Research Management techniques".

Gopal Krishna Shyam and Sunil Kumar S Manvi [2], "provided a study on RM for IaaS in cloud computing and focuses on problems in RM include adaptation, allocation, provisioning, discovery, modeling, estimation, requirement mapping and brokering. This paper focuses on some of the important resource management techniques such as resource allocation, resource provisioning, resource adaptation, and resource mapping. It emphasizes an extensive survey of these techniques for IaaS in cloud computing, and also provided the open challenges for moreover research".

Manar Bani Issa, Mahmoud Al Ayyoub, Yaser Jararweh, Mustafa Daraghmeh, and Mohammad A. Alsmirat [6], "focused on energy efficiency parameter during RM and proposed a host overloading prediction and detection algorithm based on median absolute derivation and logistic regression model. The proposed algorithm is scalable and predicts if a host overloaded or not lead to an extended performance of the RM process in the data centers. It was evaluated and compared with the five other different host prediction algorithm using CloudSim".

III. RESOURCE MANAGEMENT TECHNIQUES

Energy-aware RM techniques:- The main issues needed to be focused in cloud computing system is Energy efficiency in cloud RM. Various algorithms like Modified Best-Fit Decreasing (MBFD) [8] which is basically based on Best Fit Decreasing (BFD) [7] algorithm. In this algorithm, all Virtual Machines sorts their current CPU utilizations in decreasing order, and allocate each Virtual Machine to the host that provides the slighter increase of power consumption due to this allocation. Other algorithm, Power and

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 2 FEB 2023

Computing Capacity-Aware Best Fit Decreasing algorithm (PCA-BFD) [9] that focuses on the emphasized issue by allocating Virtual Machines to a server which has the maximum computational capacity.

SLA-aware RM techniques:- An agreement between a cloud service provider and a customer that ensures a minimum level of service is maintained according to Cloud SLA. It guarantees levels of availability, reliability and responsiveness to applications and the systems, while also specifying when there is a service interruption, who will govern it. The cloud service providers should keep a check and avoid violations while providing services to the customers. Various algorithms for SLA-aware like capacity allocation algorithms [10], task-oriented resource allocation technique [11] is proposed.

Market-oriented RM techniques:- Providers will need to consider and meet different QoS parameter of each individual consumer – So market oriented resource management is necessary to regulate the supply and demand cloud resources at market equilibrium. A resource allocation technique based on auction has been proposed in [12]. In proposed technique, if user wins the auction, he has to bid for the advertised resources which are only assigned.

Network load aware RM:- "The quantity of network traffic which is flowing via a network at some particular instant of time is termed as Network load". Various algorithm like LPBP[9] and CFMV[16] are the algorithms which minimizes network load along with energy consumptions.

Load-balanced RM techniques:- "The method of splitting workloads and computing properties in a cloud computing is termed as load balancing". It enables enterprise to manage application demands or workload demands by distributing resources among various computers, servers or networks. A new model which dynamically balances load among cloud resources has been proposed in [13]. An Artificial Neural Network (ANN) based load balancing technique has been proposed in [14]. In [15], Max–Min and Min–Min algorithms have been proposed for RM.

Table 1. Summary of RM Techniques [1]

	Energy Efficiency	Load-balanced	Market-oriented	SLA-aware	Network load	mobile clouds
PCABFD	√	×	×	×	×	×
MBFD	√	×	×	√	×	×
LPBP	√	×	×	×	√	×
CFMV	√	×	×	×	√	×
LBMM	×	√	×	×	×	×
SMDP	×	×	×	×	×	√
FRA	×	×	×	√	×	×

Resource management Aspects: -There are various aspects or issues which are needed to keep it in mind during the resource management which are listed in table-2:-

Table 2. Issues in Resource Management [2]

Issues in RM	
Resource Provisioning	"It is the allocation of resources of service provider to a customer."
Resource Allocation	"It is the distribution of resources of service provider economically among competing groups of people or programs."

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 2 FEB 2023

Resource requirement mapping	"It is an arrangement between resources available with the provider and the resources required by the users."
Resource adaptation	"It is the capacity or ability of that system to adjust the resources dynamically to fulfill the requirements of the user."
Resource discovery	"It is the Identification of a list of authenticated resources that is available for job submission and to choose the best among them."
Resource brokering	"It is the negotiation of the resources through an agent to ensure that the necessary resources are available at the right time to complete the objectives."
Resource estimation	"It is a close guess of the actual resources required for an application, usually with some thought or calculation involved."

IV. PERFORMANCE EVALUATION PARAMETERS

- **Overhead Associated** – "It ensures the amount of overhead involved during the load balancing algorithms. It consists of overhead due to movement of tasks and inter-processor communication". For efficient working of a load balancing technique, this metric should be minimized.
- **Throughput** – "It is basically the calculation of the number of tasks whose execution has been completed". To improve the performance of the system, it should be high.
  It can be evaluated by the formula,

  Throughput= J_{total} - $J_{remaining}$

  where J_{total} and $J_{remaining}$ are the total number of tasks (jobs) received and the number of tasks (Jobs) still in progress.
- **Performance** – "This parameter is used to identify the efficiency of the system". It has to be enhanced at a fair cost e.g. reduce response time while keeping some acceptable delays.
- **Scalability** – "It is the capability of an algorithm to execute load balancing for a system with any finite number of nodes. This metric should be improved".
- **Response Time** – "It is the amount of time taken to respond by a particular load balancing algorithm in a distributed system. It should be minimal to improve efficiency".
- **Fault Tolerance** – "It is the ability of an algorithm to perform uniform load balancing in spite of arbitrary node or link failure. Every system is expected to highly fault-tolerant".
- **Resource Utilization** – "This parameter is used to identify the utilization of resources in a system. It should be optimized for an efficient load balancing". Therefore, a utilization of resources should be calculated based on the overall resource by following equation.

$$U_x(F;t)1⁄4\ k1⁄41\ f_{xk}\ \square \text{Req CPU}_k\eth tÞCPU$$

Above equationpresents the utilization $U_x(F, t)$ of a server S_x at a specific time t. F represents the placement of VMs whereas f_{xk} indicates whether a VM V_k is hosted on the server or not. CPU_x represents the total computation capacity of S_x, and Req $CPU_k(t)$ is the amount of CPU capacity required by the V_k at a specified time.

- **VM Migration time** - "It is a resource intensive operation as it constantly requires adequate CPU cycles, system cache, memory capacity and network bandwidth". Thus, it contrary affects the performance of running applications and cannot be entirely overlooked in contemporary data centers, particularly when user SLA and critical business goals are to be met. VM migration time should be minimizedin order to increase the system performance.

  The researchers used the following formula to calculate migration time.

  $$T_{mj} = \frac{M_j}{B_j}$$

  where T_{mj} is the migration time, M_j is the amount of memory used by VM_j and B_j is the available bandwidth.
- **Energy Consumption** – "It determines the energy consumption of all the resources in the system. Load balancing helps in avoiding overheating by balancing the workload across all the nodes of a Cloud, hence reducing energy consumption".
- **SLA violation**- "It is an agreement between a user and the cloud service provider to ensure the required level of service". SLA violations [5] can be evaluated with following equation:

  SLAV= SLATAH .PDM

  where SLAV represents SLA Violation, SLATAH denotes SLAViolation Time per Active Host, and

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 2 FEB 2023

PDM denotes for Performance Degradation due to Migrations. We may use the following equation to calculate PDM and SLATAH.

SLATAH= 1/N+ nΣi=1 Tsi / Tai

where T_{si} is the time during which resources of host i were 100% utilized and N is the number of hosts, whereas, T_{ai} is the active time of host i.

V. CONCLUSION

In this survey, we present an extensive overview of the current Resource Management techniques which are designed for cloud computing environments. We also covered a detailed area on the Resource Management aspects that should be focused on during designing a Resource Management technique. Furthermore, we find that among the various techniques most of them either providing a solution for a single Resource Management metric or maximum two Resource Management metrics. Therefore, researchers must bring solutions for multi-metrics that should also be capable to tackle contradictory metrics. These metrics are basically dependent on each other that mean if we enhance the performance of one metric, the performance of other metric effects. Such metrics can be energy efficiency and SLA violations, profit maximization and SLA violations, and energy efficiency and load balancing.

REFERENCES

[1] Saad Mustafa, Babar Nazir, Amir Hayat, Atta ur Rehman Khan, Sajjad A. Madani, “Resource management in cloud computing: Taxonomy, prospects, and challenges”. Computers & Electrical Engineering, Volume 47, October 2015, Pages 186-203.

[2] SunilKumar S Manvi, Gopal Krishna Shyam, “Resource Management for Infrastructure as a Service (IaaS) in Cloud Computing: A survey”. Journal of Network and ComputerApplications, Volume 41, May 2014, Pages 424-440.

[3] A.R. Arunarani, D. Manjula, Vijayan Suguma-ran. "Task scheduling techniques in cloud compu-ting: A literature survey", Future Generation Computer Systems, 2018.

[4] Sukhpal Singh Gill, Peter Garraghan, Vlado Stankovski, Giuliano Casale,Ruppa K. Thu-lasiram, Soumya K. Ghosh, Kotagiri Ramamohanarao, Rajkumar Buyya, “Holistic re-source management for sustainable and reliable cloud computing: An innovative solution to global challenge”. Journal of Systems and Software, Volume 155, July 2019, Pages 104-129.

[5] Piotr Nawrocki, Wojciech Reszelewski, “Re-source usage optimization in Mobile Cloud Com-puting”.Computer Communications, Volume 99, 1 February 2017, Pages 1-12.

[6] Yaser Jararweh, Manar Bani Issa, Mustafa Daraghmeh, Mahmoud Al-Ayyoub, Mohammad A. Alsmirat, “Energy efficient dynamic resource management in cloud computing basedon logistic regression model and median absolute deviation”. Sustainable Computing: Informatics and Systems, Volume 19, September 2018, Pages 262-274.

[7] Frederic Nzanywayingoma & Yang Yang, “Ef-ficient resource management techniques in cloud computing environment: a review and discussion”. International Journal of Computers and Applica-tions, January 2018, doi:1-18. 10.1080/1206212X.2017.1416558.

[8] Beloglazov A, Abawajy JH, Buyya R, “En-ergy-aware resource allocation heuristics for effi-cient management of data centers for Cloud com-puting” FutureGener Comput Syst 2012;28(5):755–68.

[9] Tziritas N, Xu C-Z, Loukopoulos T, Khan SU, Yu Z, “Application-aware workload consolidation to minimize both energy consumption and network loadin cloud environments” In: 42nd IEEE international conference on parallel pro-cessing (ICPP); 2013.

[10] Ardagna D, Casolari S, Colajanni M, Panicucci B, “Dual time-scale distributed capacity allocation and load redirect algorithms for cloud systems” Journal of Parallel Distrib Comput 2012;72(6):796–808.

[11] Ergu D, Kou G, Peng Y, Shi Yong, Shi Yu, “The analytic hierarchy process: task scheduling and resource allocation in cloud computing envi-ronment’’. JSupercomput 2013;64(3):835–48.

[12] Zaman S, Grosu D, “A combinatorial auc-tion-based mechanism for dynamic VM provisioning and allocation in clouds” IEEE Trans Cloud Comput2013;1(2):129–41.

[13] Ban Y, Chen H, Wang Z, “EALARM: an ENHANCE autonomic load-aware RM for P2P key-value stores in cloud” In: 7th IEEE interna-tional symposiumservice-oriented system engi-neering; 2013.

[14] Al Sallami NM, Al Daoud A, “Load balancing with neural network” Int J Adv Comput Sci Appl (IJACSA) 2013;4(10):138–45.

[15] Kokilavani T, George Amalarethinam DI, “Load balanced min–min algorithm for static me-ta-task scheduling in grid computing” Int J Com-put Appl2011;20(2).

[16] He L, Zou D, Zhang Z, Chen C, Jin H, Jarvis SA. Developing resource consolidation frame-works for moldable virtual machines in clouds. Future GenerComput Syst 2014; 68–81.