DOI: 10.56452/2021SP-8-010

ISSN: 0974-5823 Vol. 6 (Special Issue 3, November 2021)

International Journal of Mechanical Engineering

CloudSim: Cloud Computing Environment Modelling and Simulation as well as Resource Provisioning Algorithm Assessment

Kiran Bala[1], Sanjay Kumar[1] and Arti Rani[2]

[1]Department of Computer Science & Engineering, Chandigarh Engineering College, Jhanjeri, Mohali-140307, Punjab, India

[2]Department of Agriculture, Chandigarh School of Business, Jhanjeri, Mohali-140307, Punjab, India

Abstract: The lifelong ambition of computing as a utility, cloud computing, has the power to transform a significant portion of the IT sector, increasing the appeal of software as a service and influencing the development and selection of IT hardware. New hardware investments and labor costs for operating Internet services are no longer necessary for developers of innovative new invention concepts. They need not be concerned about over-or under-provisioning for a service whose demand turns out to be much higher or lower than expected, squandering expensive resources and maybe losing out on consumers and income.

Keywords: Cloud Computing, Virtualization, SaaS, PaaS, IaaS, VM scheduling

Introduction

Internet-based technologies are used by cloud computing to supply IT-enabled services to any desired users. With cloud computing, we can access anything from anywhere from any computer without having to worry about things like management, pricing, storage space, infrastructure, and other things. Cloud computing is the term for the Internet-based delivery of computing services. Instead of preserving data on your own hard drive or updating software to meet your needs, you utilize services in any location, to maintain your data or even use its applications. Utilizing software and hardwarethat is controlled by external parties at different locations is permitted by cloud services for both individuals and corporations [4]. Social networking sites, webmail, online business tools, online file storage, skype, mails, peer-to-peer networks, and many more are examples of cloud services. Information and computer resources can be accessed via the cloud computing approach from any location with a network connection. Cloud computing provides access to a pool of resources that are shared, such as machines, storage systems, applications, computing power, networks, and particular business and user applications. [1]. Through the use of virtualization technology, clients can dynamically scale their rented virtual machines (VMs) to include as many machines as they require by making a credit card payment. Government agencies, academic institutions, and corporate groups can reduce operating costs with the aid in the cloud computing. The term "cloud computing" has been defined as followed by the U.S. National Institute of Standards and Technology (NIST) [2].

“A model called "cloud computing" makes it possible to quickly create and deploy a pool of shared, adaptable computational power (such as networks, servers, storage, applications, and services) via a network while requiring little management labour or service provider interaction. This cloud model encourages availability and consists of three service models, four deployment models, and five important qualities.”

A group of interconnected, virtualized computers that are continuously provisioned [3] and available as one or even more unified computing resources in accordance with service-level agreements negotiated between both the service provider andcustomers make up the type of parallel and distributed system known as cloud computing. Microsoft Azure, Amazon EC2, Google App Engine, and Aneka are a few examples of new cloud

computing infrastructures and platforms.The properties of all these technologies are often present in cloud computing [5].

Here is a brief description of these innovations.

Literature survey

Kosar et al. (2004) Task scheduling is a difficult problem as well as being crucial to scientific operations. It has already been studied in relation to traditional distributed computing systems. The Grid scheduler that ensures fault-tolerant queueing, monitoring, programming, and management of task scheduling activities[3].

Rafique et al. (2011) They examined different options for system design and capacity-aware job scheduling for large-scale data processing on circulating systems that use accelerators. The runtime support for exploiting various computing accelerators through runtime workload adaption and for adaptively transferring MapReduce workloads to accelerators in virtualized execution settings enhances the MapReduce programming model. The processing costs and transmission times between data centres on the Internet, however, are not the emphasis of any of them. As the importance of cloud computing has grown, new data management systems, such Google's GFS (Google File System) and Hadoop, have been developed[4].

Bossche et al. (2011)Because users must pay rent for resources according to resource utilisation, most people use cloud computing resources rather than their own computers. The scheduling of computing applications, which must be economical, is done using an algorithm. This method is also used for virtual machine data transfer times[5].

Li et al. (2012) the best load balancing strategy is employed to address the issue. This method also cuts down on the amount of time needed for computation when it is used[5].

Breitgand et al. (2011)In a federated cloud scenario, one cloud can hire the other cloud to fulfil user demands. The major goals are cloud consolidation and load balancing, as well as cost reduction. The two methods, which are based on avaricious standards, are employed6].

Tordsson et al. (2012)Some services are delivered using both single and many clouds. Performance through a multi-cloud setup is superior to performance through a single cloud. The high throughput can be provided by these clouds[7].

Grid computing

A network of computers is used in grid computing to get powerful supercomputing-style computing resources. Large-scale, complicated computing activities can be carried out with this computer network. These computer networks may exist in several locations when using grid computing. Folding@Home is a well-known Grid Computing project. The project involves solving a challenging scientific problem using the idle computing capacity of thousands of computers. Understanding protein folding, misfolding, and associated disorders is the project's main objective.

Virtualization

Between the operating system and the hardware, virtualization adds a layer. In the 1960s, mainframes began to accommodate numerous users who used virtual machines. For each user, these virtual computer reproduced how an operating system would behave. In 1999, VMWare introduced VMware Workstation, a solution that enables the use of various operating systems on personal computers. The core of cloud computing is virtualization. Users have access servers or storage using virtualization without being aware of individual server or storage information. By utilizing the necessary resources, the virtualization will carry out user requests for computing resources. Server usage in datacentres is typically as little as 10%. The usage of servers can be greatly increased with virtualization[1].

Utility Computing

A "pay-per-use" paradigm for utilizing computing services is defined by utility computing. The way computing resources are billed is analogous to the way traditional utilities such electricity are billed. The upfront cost is very low when we buy electricity from a vendor. The customer receives a bill from the electricity provider based on their usage (typically monthly). Similar billing protocols are used in utility computing. Different billing structures are being investigated. Among the most typical are:

• Billing for each user. For instance, if a company with 100 employees utilizes Microsoft Live or Google's Gmail as its internal email system, with the company's email stored on cloud servers, The business might be charged on a per-user basis by Google or Microsoft.

• Gigabyte-based charging. Amazon may charge a firm for the disc space utilized if the company uses Amazon to host its data in the cloud.

• Billing per day or hour. For instance, A customer could pay the use of virtual servers according on the number of hours they use them.

Model for cloud computing

All A collection of IT services known as "cloud computing" are offered Providing an user across a channel on a leased premise with the ability to scale their service requirements up or down. Typically, a third party provider who controls the infrastructure provides cloud computing services. An innovative business model for adopting IT services by enterprises is provided by cloud computing. The Cloud service model and the Cloud deployment model are the two fundamental cloud models that are covered [6].

Cloud Service Model

In a cloud computing environment, enormously flexible IT-related skills are made available as a web - based service to a large range of external consumers. This phrase accurately captures the various aspects of the Cloud Computing paradigm that exist at various infrastructure layers. IaaS, PaaS, and SaaS are the three services that make up cloud computing. Different utility services are offered via cloud computing.

a) *IaaS (Infrastructure as a service) model:* This paradigm's main principle is virtualization, in which users have virtual desktops. and use cloud service provider-supplied network, storage, virtualized servers, routers, and other resources. Its cost of usage is determined by the number of CPU hours, data GBs stored, network bandwidth used, network infrastructure required, and value-added services employed, such as monitoring and auto-scaling. Examples are the AmazonS3 and Amazon EBS storage services. Computing services, including Layered Technology and Amazon EC2. Everything as a service is another name for this business model.
b) Amazon EC2: Amazon Elastic Cloud is an online service that provides scalable processing capacity in the cloud. Businesses may easily access and configure capacity with Amazon EC2 thanks to its straightforward web service interface. It gives users access to computing resources and enables use of Amazon's computing infrastructure by enterprises. As computing needs change, Amazon EC2 enables rapid capacity scaling up and down by cutting the duration required to buy and boot additional server instances to minutes. With Amazon EC2, you may alter the economics of computing by only paying for the capacity that you really utilise.
c) *PaaS (Platform as a service) mode-l*In order to enable the direct deployment of applicationlevel assets or web apps, it refers to the environment that offers the runtime environment, software deployment framework, and component on pay. A platform for the development, testing, and deployment of software is known as PaaS. It implies that a PaaS can handle the whole life cycle of software. Application developers, testers, deployers, and administrators are the focus of this service model. Examples include Amazon EC2, Salesforce.com, Jelastic.com, Microsoft Azure, IBM Smart Cloud, Google App Engine (GAE), and others.
d) Salesforce.com: The Salesforce.com provides technology is called Force.com. Force.com's cutting-edge Visual Force technology offers the first user interface-as-a-service, enabling users, programmers, and independent software providers to build any application for any consumer, anywhere. The Force.com platform offers a global services and infrastructure for storage, logical, workflow, integration, ui, and application exchange.

***SaaS (Software as a service)*:** Through such a paradigm for providing services, end users directly access network-based software application services on an as-needed basis. For instance, Google serves as the

supplier while we, the users, are the customers of Gmail, a SaaS. Arial System's and Op Source's billing services are two further well-known instances of PaaS.
For Google Apps, the major features are available:

Gmail With the help of Gmail, companies can now send, receive, and store email messages with ease, retaining a full keeping track of client interactions to improve sales execution and boost client satisfaction.

GDocsChrome Docs Within your sales organization, marketing team, or support team, For instant collaboration, develop, edit, and share online Google Documents, Google Spreadsheet applications, and Google Presentations.

Google Talk on Google Instantly communicate with coworkers or clients from Salesforce, with the opportunity to add Google Talk discussions to client or prospect data.

Google Calendar displays Salesforce's sales tasks and marketing initiatives. This application, created by Appirio, is an illustration of a brand-new class of Salesforce for Google Apps partner extensions.

Cloud Deployment Model
There are four main deployment types for cloud computing that are provided to service users.

1) *Public cloud/external cloud:* This architecture permits open or public access to the cloud environment. Off-premises, public clouds allow a variety of businesses to provide services to customers by obtaining them from other sources.
2) *Private cloud/internal cloud:* In this approach, the term "on-premise cloud" referred to cloud infrastructure and services that are managed or owned on-site by a company. In other words, a private cloud was created particularly to offer services to a company so that security and privacy could be maintained.

3) Hybrid cloud/virtual private cloud model: The cloud computing environment is hosted and managed by a third party (off-premises) under this hybrid approach, although An enterprise can only use specific specialised resources in private.

4) *Community model:* It permits the use of aenvironment for cloud computingthat is shared or operated by severalaffiliated businesses.

Virtualization
A digital representation of something(opposed to the real thing), such as anetwork resources, an operating system, a server, or a storage device can be created by virtualization. The primary component of cloud computing is a market-oriented distributed computing system made up of a network of connected virtualized machines that can dynamically present themselves as one or more unified computing resources depending on the terms of the Service Level Agreement (SLA) that is negotiated between the service provider and the customer[2].

VM scheduling
Processes or tasks are planned according to the provided requirements and the employed algorithm in a balanced scenario called scheduling. According to the requirements met by the requested resources, VM scheduling algorithms are used in cloud computing to schedule VM requests to the Physical Machines (PM) of the specific Data Center (DC) (i.e. RAM, Memory, Bandwidth etc). There are numerous cloud service providers with varying capacities of data centres and physical machines on the market today. Leading cloud providers in 2013 included SalesForce, Google Apps, Amazon, MS Office and Microsoft Azure, Oracle and others. In general, the three stages of the scheduling algorithm are as follows: [11]

Execution of the task
• The task is assigned by the user for cloud execution.
• The cloud coordinator receives the assignment (CC).
• The work is forwarded by the cloud coordinator to the datacenters (DC).

• There are numerous unfixed hosts in datacenters that house a pool of virtual machines (VM).
• These hosts may be modified or eliminated as necessary [2].

Allocation of resources for execution of task:

• VMM sends the task requirements to the resource provisioner to request resources.
Resource provisioner checks with Resource Owner to see if resources are available.
•If the resources are available, the resource owner gives the resource provisioner access authority to use the resources.

Virtual Machine Migration

Think of a server that has a hypervisor, numerous virtual machines, each of which is running an OS and apps. Application availability will be greatly impacted If you need to restart the software components after shutting down the server for maintenance (such as adding more memory to the server). VM migration enables you to transfer a full virtual machine (complete with its installed operating system and apps) from one to another machine while maintaining the VM's functionality on the new machine.
Because physical servers may be taken down for maintenance with little disruption to running applications, virtualized environments have this benefit over physical ones. After the VM on the source machine has been put to sleep, You can perform this migration by launching the VM on the target system, moving its supporting data to the destination machine, and relocating the VM itself. To minimize downtime, you can do this migration while the virtual machine is running and continue its operation on the target computer once all the state has been migrated. [14]

Problem definition and proposed work

If we concentrated on the pull model and the work done in the aforementioned literature, we could see that these scheduling algorithms were pull-focused. As a result, present methods require a stage that heavily relies on computational resources. They consistently disregard where the data is located. This results in significant bandwidth usage costs. When data is really large, locality cannot be ignored. This necessitates a different approach where both the data host and compute host should be chosen, as opposed to the data host being chosen first and the compute host being chosen afterwards, or vice versa.

The workflow challenges with distribution and planning in cloud systems are addressed using three complimentary bi-criteria techniques. Moreover, we have suggested two lower bounds for each taken criterion in order to evaluate the quality of the solutions produced by our methods.

These methods, in contrast to earlier efforts, simultaneously consider two criteria that are in conflict:
• Cost of execution
• Time for execution.

Additionally, they give consumers greater flexibility to gauge their preferences and select a desirable schedule from the successful outcomes. More specifically, the first strategy concentrated on the cost associated with employing a particular set of resources, whereas the second aims to reduce the overall execution time. The two first methods for choosing solely the Pareto solutions form the basis of the third method. Although the results in this research are quite positive, they also raise fresh and intriguing issues. For this work, we therefore concentrate on a number of viewpoints. We are currently expanding the suggested methods for business process workflows to take into account workflow compositions and patterns. [8].

Problems Identified and Our Suggestion

Our strategy is to bring together data-intensive operations in order to:
• Examine the financial benefit of a comprehensive and beneficial relationship between the set of resources that have been purchased and used to carry out a task.
• Establish closer ties between a container management service and the underlying physical infrastructure.
• To put in place proactive job monitoring programmes that assess job status using historical pattern inference.

This work's primary goal is to offer a concept for the upcoming cloud computing generation that is built on a virtual infrastructure container that organizes resources holistically.The suggested framework effectively distributes tasks among the resources provisioned as part of the container, automates the synchronization of activities within the container, and optimizes resource provisioning.

Proposed algorithm

The workflow algorithm is as follows:
Step 1: Randomly initialize the request location vector with a request dimension equal to the particular job size.
Step 2: Determine the fitness value for each particle by breaking it out into various tasks.
Step 3: Directly link tasks that depend on one another.
Step 4: Picking the top particle out of all the others to be the greatest.
Step 5. Update each virtual machine with the latency between them.
Step 6:Map the VMs to the tasks according to latency
Step 7:Execute the workload to the VMs [9].

Conclusion and future work

These different requests relating to a single workflow must be handled in order to fully utilise the resources. In this study, we examined the reliability of the workflows and connected them to the overall time spent switching between virtual machines. In this study, we devised a method for transferring massive amounts of data in applications that require it. We put the work into practice to speed up task reaction time. Depending on their dependability, the jobs are broken down into subtasks and a workflow is developed. According to the latency between virtual computers, the tasks are distributed among them. The suggested technique reduces response time because it is based on transfer time. and increases the number of requests processed in a particular and raises the quantity of requests handled in a specific period of time. The existing work does not offer this capability because it takes longer to respond to queries and can only handle a limited number of them at one time.
Future attempts at tying workflow management to virtual machine costs and workflow requests will be made. Users will benefit from this by saving money. The workflow in our work can be the same for multiple tasks, but in the future, we can try a different workflow for each activity.

References

[1] Mladen A. Vouk, “Cloud Computing – Issues, Research and Implementations”, Journal of Computing and Information Technology, pp. 235–246, 2008.
[2] A. Beloglazov, J. Abawajy, R.Buyya, “Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing”, Future Generation Computer Systems, pp. 755-768, 2012.
[3] R. Buyya, J. Broberg, A. Goscinski, “Cloud Computing principles and paradigms”, A John Wiley and sons, Inc. Publication, 2011.
[4] R. Buyya, S. Pandey, “Data Intensive Distributed computing: Challenges and solutions for large-scale information system”, Information science reference, 2012.
[5] K. Bessai, S. Youcef, A. Oulamara, C. Godart, S. Nurcan, “ Bi-Criteria workflow tasks allocation and scheduling in Cloud computing environments”, IEEE 5th international conference on cloud computing, pp. 638-645, 2012.
[6] KahinaBessai, Samir Youcef, Ammar Oulamara, Claude Godart and SelminNurcan, “Multi-objective resources allocation approaches for workowapplications in Cloud environments”, OnTheMove Federated Conferences & Workshops, 2012.
[7] KahinaBessai, Samir Youcef, Ammar Oulamara, Claude Godart and SelminNurcan, “Resources allocation and scheduling approaches for business process applications in Cloud contexts”, 2012.
[8] Van den Bossche, Belgium Vanmechelen, K. Broeckhove, “Cost-Efficient Scheduling Heuristics for Deadline Constrained Workloads on Hybrid Clouds”, IEEE Third International Conference on Cloud Computing Technology and Science, pp. 320-327, 2011.
[9] D. Breitgand, A. Maraschini, J. Tordsson, “Policy driven service placement optimization in federated

cloud", IBM, pp. 1-10, 2011.
[10] A. Chervenak, D. Smith, W. Chen, E. Deelman, "Integrating policy with scientific workflow management for data intensive applications", 2012.
[11] J.M. Cope, N. Trebon, H.M. Tufo, P. Beckman, "Robust data placement in urgent computing environments", IEEE International Symposium on Parallel &Distributed Processing, IPDPS, pp. 1–13, 2009.
[12] R. Calheiros, R. Ranjan, A. Beloglazov, C. Rose, RBuyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", Wiley online library, pp. 23-50, 2011.
[13] E. Deelman, A. Chervenak, "Data management challenges of data intensive scientific workflow", 2008.
[14] Kumar, Sanjay, Ashish Negi, J. N. Singh, and Amit Gaurav. "Brain tumor segmentation and classification using MRI images via fully convolution neural networks." In *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 1178-1181. IEEE, 2018.
[15] Kumar, Sanjay, Ashish Negi, J. N. Singh, and Himanshu Verma. "A deep learning for brain tumor MRI images semantic segmentation using FCN." In *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1-4. IEEE, 2018.
[16] Kumar, Sanjay, Syed Akbar Abbas Jafri, NishitArun Nigam, Nakshatra Gupta, Gagan Gupta, and S. K. Singh. "A new user identity based authentication, using security and distributed for cloud computing." In *IOP Conference Series: Materials Science and Engineering*, vol. 748, no. 1, p. 012026. IOP Publishing, 2020.
[17] Kumar, Sanjay, Sanjeev Kumar Singh, Yash Gupta, Vasu Bajpai, and Piyush Tiwari. "Brain Tumour MRI Image Segmentation and Classification Using FCNN." In *Soft Computing for Security Applications*, pp. 323-329. Springer, Singapore, 2022.
[18] Kumar, Sanjay, Naresh Kumar, Inderpreet Kaur Rishabh, and Vivek Keshari. "Automated brain tumour detection using deep learning via convolution neural networks (CNN)." *Int. J. Cur. Res. Rev* 13, no. 02 (2021): 148.
[19] Kumar, Sanjay, J. N. Singh, and Naresh Kumar. "Brain tumor detection & segmented classification of CT-scan and MRI image." In *Smart Computing*, pp. 290-297. CRC Press, 2021.