IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, May 2023

Integration of Resource Assignment and Migration for Effective Resource Optimization in Cloud Utilization

Aruna. A[1], Aarthy S[2], Asha Nandhini G[3], Deva Dharshini J K[4]

Assistant Professor, Department of Computer Science[1]

Students, Department of Computer Science[2,3,4]

Dhanalakshmi College of Engineering, Chennai, India

Abstract*: Numerous processing opportunities and heterogeneous resources are available through cloud computing, which also satisfies the needs of numerous applications at different levels. Resource allocation is one of the most important aspects of cloud computing. The performance of the entire cloud environment is directly impacted by its efficiency. Thus, in cloud computing, resource allocation, and management are crucial. Resource allocation is a method of distributing the available resources—such as the CPU, RAM, storage, and network bandwidth—among users in cloud data centers in a way that promotes efficient use of those resources, provider profitability, and user pleasure.*

Keywords: Cloud computing, resource migration, cache memory, Task scheduling, data centers, Cloud service provider

I. INTRODUCTION

In recent years, businesses, governments, and academic institutions have utilized cloud computing more and more. It is used in Cloud Data Centres (CDCs), which are facilities for managing numerous large-scale infrastructures, many of which contain millions of servers and cooling systems. The infrastructure resources in CDCs are pooled to support numerous applications running at once and provide services to clients anywhere in the world. Tens of megawatts of energy are typically required by each CDC to power and cool tens of thousands of servers. Each application is often deployed in numerous CDCs situated in various geographical areas to provide high availability and low latency. Each CDC is also connected to a number of Internet service providers (ISPs), which distribute enormous amounts of data across the spread of CDCs and millions of customers while also taking performance and cost into account. It is demonstrated that the majority of a CDC provider's operational costs are attributable to ISP bandwidth and energy costs. Recent years have seen a sharp rise in the variety and quantity of applications. According to recent estimates, the energy used by data centers in the United States in 2017 was around 78 billion kWh or 2.9% of all the energy used in the country. It is anticipated that by 2020, this percentage will have reached 4%. In 2017, burning renewable fossil fuels like coal, oil, and natural gas produced more than 80% of the energy used in the United States. The worldwide ecology would suffer irreparable damage and pollution as a result. In order to mitigate the environmental damage brought on by the use of fossil fuels, an increasing number of CDCs, including Microsoft and Amazon, deploy renewable energy facilities and transition to distributed green CDCs (DGCDCs). The energy cost of the DGCDCs is increasing dramatically along with the tasks of applications. Studies on energy optimization issues have been conducted by both businesses and academic institutions. However, before users' tasks can reach back-end DGCDCs, they must first pass through the wide-area network, which includes numerous ISPs. For instance, Google's wide-area network offers a variety of services to users all over the world, including video, search, and mail. Due to users' duties and reaction data being communicated between users and DGCDCs, DGCDCs also need to pay money to ISPs. Currently, typical DGCDCs transmit more than a petabyte of data every day, which makes their ISP bandwidth costs extremely high. Additionally, a service-level agreement (SLA) established between users and a DGCDC provider determines the bandwidth price of each ISP. As a result, some ISPs have much higher bandwidth prices than others.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

Caches, computing cores, and the Network-on-Chip subsystem are all taken into account in a workable system model for Network-on-Chip designs that successfully capture energy usage. However, it only takes into account data center energy consumption optimization. An energy and cost-aware task scheduling approach that enables mobile devices to schedule some work to cloud resources is proposed by Nir et al. The best solution can drastically lower the cloud provider's overall cost. It does not, however, take into account the energy generated by renewable resources. A task scheduling model that Wang et al. suggest specifies the performance requirements of tasks with the lowest frequency. In order to save energy, it distributes incoming jobs among running servers and modifies the server cores' execution frequencies.

Users' tasks must first pass through the wide-area network, which is made up of several ISPs, in order for them to reach back-end DGCDCs. For instance, Google's wide-area network provides customers all over the world with a range of services like video, search, and mail. Due to the obligations of users and the exchange of reaction data between users and DGCDCs, DGCDCs also have financial obligations to ISPs.At the moment, typical DGCDCs transmit more than a petabyte of data per day, which results in extraordinarily high ISP bandwidth charges. The bandwidth cost of each ISP is also determined by a service-level agreement (SLA) made between users and a DGCDC provider. Because of this, some ISPs charge substantially greater rates for bandwidth than others.

II. RELATED WORKS

The provisioning of virtual machines (VMs) in the cloud has been studied using completely unanticipated reading and application marking. In this part, we'll generally examine the tightly related VM provisioning tools used by public clouds, private clouds, assets-driven clouds, and heterogeneous clouds, as well as tools that start-up VMs for demands with sporadic request caps.

Closeout-based models for VM provisioning and allocation have been introduced by cloud providers, allowing clients to submit offers for the VMs they have specified. We will generally look at the best components for the situation, which showed that the cloud provider setups VMs upheld the successful clients' requests and determined their payments. By lying about the groups of VM events they have mentioned and their values, we will generally demonstrate that the customers lack incentives to manage the framework.

III. EXISTING SYSTEM

The infrastructure resources in distributed green cloud data centers (DGCDCs) are shared by multiple heterogeneous applications to provide flexible services to global users in a high-performance and low-cost way, in the existing system.

Drawbacks

Congestion occurs during job allocation. More power consumption, due to utilization of multiple physical machines, at times in case of the same job. Waiting time is increased when job allocation is not done efficiently.

IV. PROPOSED SYSTEM

Here, we have deployed two types of systems. 1. Hot Machines can handle the current job. 2. Warm Machines are kept idle state until the job is assigned. We deploy three Virtual servers for every machine. 1st Job is assigned to the Hot machine of the 1st Virtual machine. Similarly, the following jobs are assigned to other VMs. Now jobs are assigned to the Warm machines once all the VMs of the Hot category have occupied with the jobs. Automatic migration of the job is implemented, so as to transfer the load to the Hot VM from Warm VM once it has completed the job. We have also implemented a cache mechanism.

Merits

- Avoid congestion and replicate the request
- Less power consumption
- Waiting time is decreased
- Reliable
- High data transmission rate

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

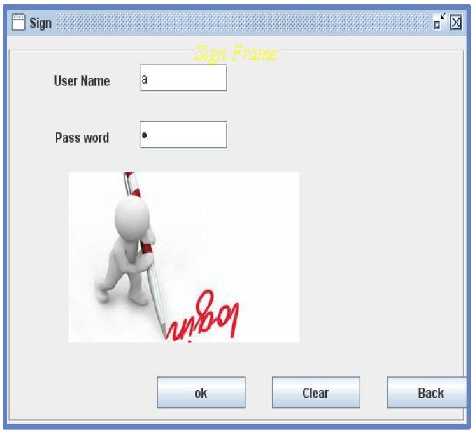
V. MODULE DESCRIPTION

By enabling the concurrent development of various system components, a modular design minimizes complexity, facilitates modification (a crucial component of software maintainability), and makes implementation simpler. Software architecture, in which software is divided into discretely named and addressable components known as modules that are integrated to satisfy problem requirements, is an example of software that effectively embraces modularity.

User registration:

Create a user application that enables access to data on the cloud service provider's server by the user. The first step is for the User to register, and only then will they be given access to the Network. The User must connect to their account after creating one to request the Job from the Cloud Service Provider (CSP). CSP processes the requested Job and responds based on the user's request. All user information will be kept in the cloud service provider's database.

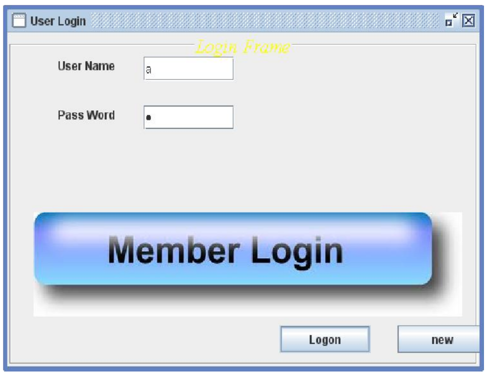
In this project, we'll create the User Interface Frame to Use Network Coding to Communicate with the Cloud Server Using Java/.Net. By submitting a request to a cloud server provider, the user can only access the required data after CSP has verified their identity.



Cloud Server Deployment:

Massive amounts of data are stored by cloud service providers. Additionally, the cloud service provider keeps track of user data that is used to verify users when they login to their accounts. The user data will be kept in the cloud service provider's database. Additionally, the Cloud Server routes the user's job request to the Resource Assigning Module for processing. The Resource Assigning Module will handle the users' requests. The Cloud Server will create a connection between them in order to communicate with the Client and the other Cloud Network modules. We're going to make a user interface frame with this purpose in mind. Additionally, the Cloud Service Provider will FIFO-send the User Job request to the Resource Assign Module.

Intermediate Server Deployment:



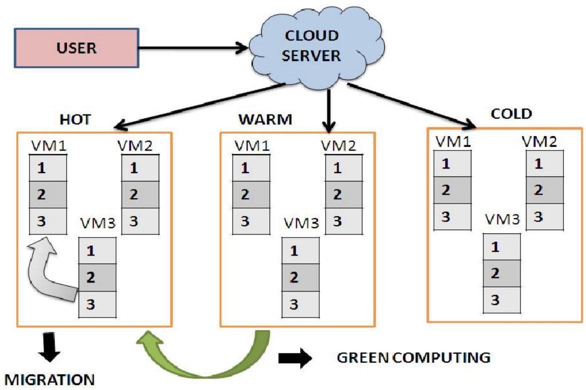
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

The User Requested Job will be processed successfully by the Job Processing Scheme when the Intermediate Server is implemented, and the Cloud Server's resources will be efficiently maintained. so that we can conserve resource energy when a job isn't completed.

Green Computing Setup:

The phrase "green computing" refers to resource-efficient computing. It also goes by the name Green IT. We will handle the user-requested Job in this Module. The user-requested job is forwarded to the cloud server's RAM. Three different types of physical servers will be present in the RAM., namely

1. HOT Server
2. WARM Server and
3. COLD Server.



These Physical Servers will contain the 'n' number of virtual Servers to process the User requested Job. Such that, the Job will be efficiently processed.

Migration of Virtual Server:

It is possible to move a process from one virtual machine to another without losing any data by utilizing the migration server, a module that is used to move jobs from one virtual server to another server. This balances the workload on the server and reduces energy consumption.

Cache Server Implementation:

We are adding a cache memory to this project as a change so that the user's requested job will be saved for a while. The server of the Cloud Service Provider (CSP) will examine its cache memory first if another User requests the same Job from it. so that we can shorten the time needed to process jobs. The server will give the user the requested data right away if it is presented. If the requested data is not already in the cache memory, the server will handle the user's request by moving the desired data to RAM.

VI. OUTPUT

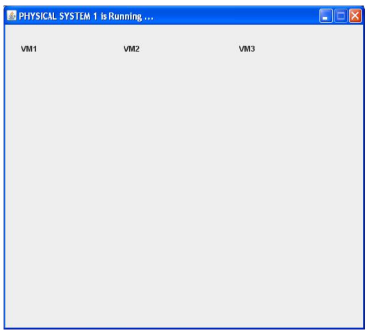


Fig 1.1 physical machine running

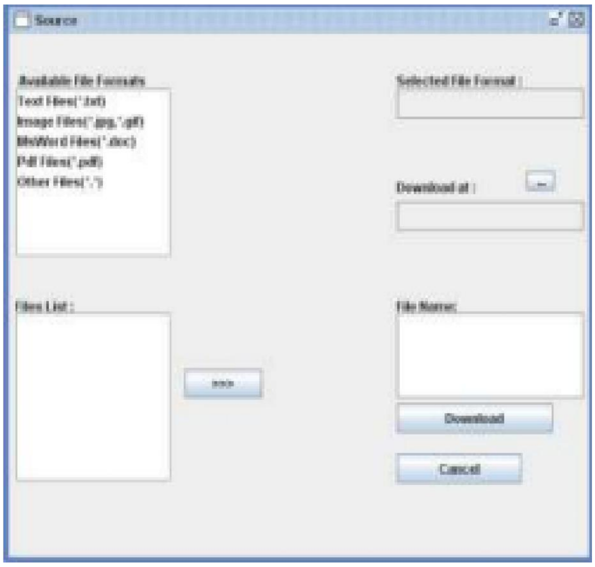


Fig 1.2 request from user

VII. CONCLUSION

If there is different and distinct burstiness in a highly consolidated computing cloud, the VM performance is susceptible to degradation in the absence of an appropriate VM placement strategy. More PMs need to be triggered in order to address this issue, consuming more energy. We have suggested setting aside a specific amount of resources for each PM to create a queuing system that can handle burst workloads in order to balance performance and energy costs. The measurement of the number of reserved resources is not an easy task. In this research, we present a two-state Markov chain-based burstiness-aware server consolidation approach. We demonstrate that the suggested algorithm can guarantee this performance restriction by using a probabilistic performance constraint.

VIII. FUTURE WORK

Further data security can be implemented through different algorithms such as AES, which includes encryption also, and blockchain integration will ensure further security. Data chunking and multiple cloud integration (dropbox and google cloud) ensure data security and multiple usages

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M.Zaharia, "A view of cloud computing," Commun. ACM, vol. 53, no. 4, pp. 50–58, 2010.

[2] M.-H. Oh, S.-W. Kim, D.-W. Kim, and S.-W. Kim, "Method and architecture for virtual desktop service," U.S. Patent 20 130 007 737, 2013.

[3] M. Marzolla, O. Babaoglu, and F. Panzieri, "Server consolidation in clouds through gossiping," in Proc. IEEE Int. Symp.World Wireless, Mobile Multimedia Netw., pp. 1–6.

[4] W.Vogels,"Beyondserver consolidation," ACM Queue, vol. 6, no. 1, pp. 20–26, 2008.

[5] N. Bobroff, A. Kochut, and K. Beaty, "Dynamic placement of virtual machines for managing SLA violations," in Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manag., 2007, pp. 119–128.

[6] W.-T. Su and S.-M. Wu, "Node capability aware resource provisioning in a heterogeneous cloud," in 2012 1st IEEE International Conference on Communications in China (ICCC). IEEE, 2012, pp. 46–50.