IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# A Review of: Cloud Computing Distributed System using Load Balancing

**Shilpa Rani 1**

**Assistant Manager, Department of School Education , HSSPP, Panchkula**

## Abstract

Cloud computing is a rapidly growing paradigm that allows all IT needs, such as storage, compute, and applications such as office and ERP, to be outsourced over the Internet. Virtualization, utility computing, pay-as-you-go, no capital investment, elasticity, scalability, provisioning on demand, and IT outsourcing are all examples of cloud computing technologies and ideas. Cloud computing is characterized as dynamically scalable shared resources that are totally available through a network, with users only paying for what they use and having the ability to share resources internally or with other customers. Cloud computing provides infrastructure, platform, and software (applications) as services, which are made available to users as pay-as-you-go subscription-based services. Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) are the terms used in the industry to describe these services. Cloud computing benefits IT organizations by relieving them of the low-level burden of setting up fundamental hardware and software infrastructures, allowing them to focus on innovation and adding value to their services.

## Introduction

## 1.1 CLOUD COMPUTING

A cloud is a type of internet or network. To put it another way, a cloud is anything that existing at a faraway area. A vast number of computers are connected via a communication network such as the Internet in cloud computing [7]. Cloud computing refers to the use of the internet to alter, configure, and access programmes. It provides online data storage, application development, and infrastructure. Cloud computing gives us the ability to access apps as utilities through the internet. It enables us to design, configure, and personalise apps through the internet. Cloud computing can deliver services across a network, such as a private network or a public network, such as a LAN, WAN, or VPN. Web conferencing, email, and customer relationship management (CRM) are all cloud-based applications.

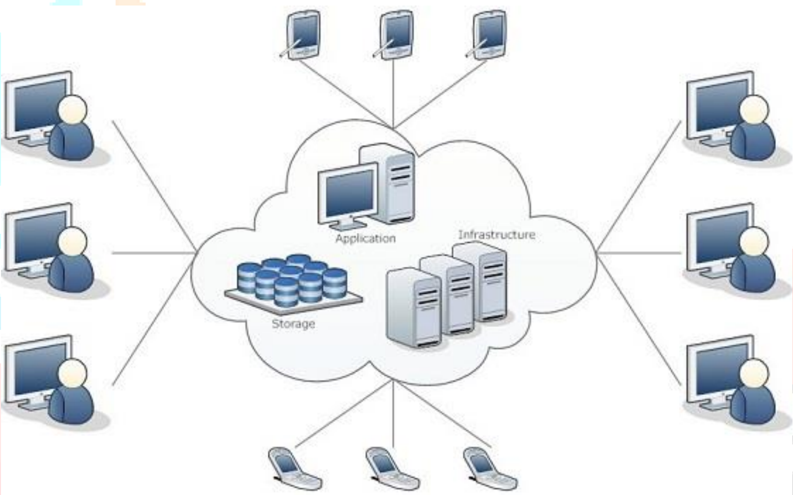


Fig: 1.1 Cloud Computing Infrastructures

## 1.2 CLOUD COMPUTING ARCHITECTURE

The Cloud Computing architecture is made up of a number of loosely linked components. The cloud architecture is divided into two parts:

- Front End
- Back End

### 1.2.1 FRONT END

Front End is the client part of cloud computing system. It comprises of interfaces and applications that are required to access the cloud computing platforms. Example: - Web Browser

### 1.2.2 BACK END

The cloud is the back end. It includes all of the resources necessary to deliver cloud computing services. It includes large amounts of data storage, security mechanisms, virtual machines, deployment methods, services, and servers, among other things.

The back end is responsible for providing traffic control and a built-in security mechanism. Each end is connected through a network, which is generally the Internet.

The figure below depicts a graphical representation of cloud computing architecture [8]:

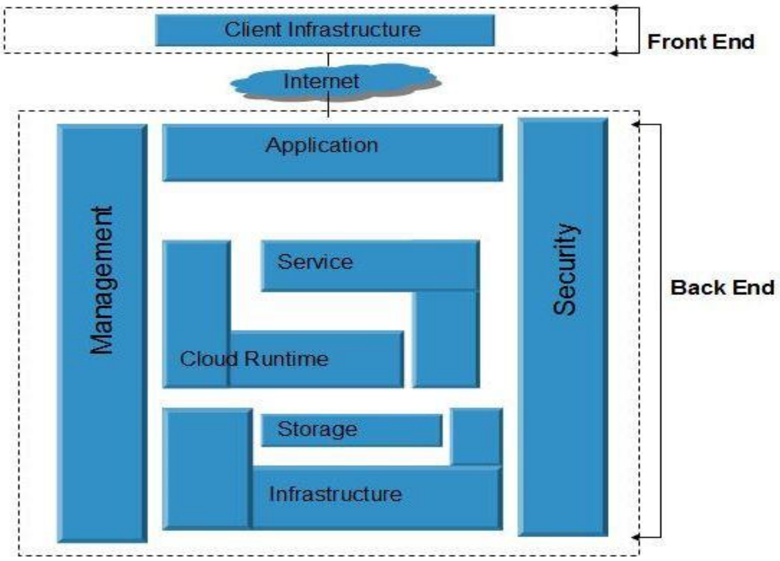


**Fig: 1.2 Cloud Computing Architecture**

## 1.3 CLOUD INFRASTRUCTURE COMPONENTS

Cloud infrastructure [8] consists of storage, servers, network, management software, and deployment software and platform virtualization.

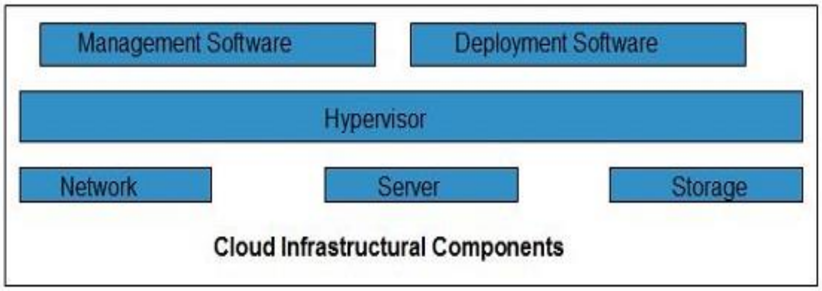


**Fig: 1.3 Cloud Infrastructure Components**

- **Hypervisor:** Hypervisor is a firmware or low level program that acts as a Virtual Machine Manager (VMM). VMM allows sharing the single physical instance of cloud resources between several tenants.
- **Management Software:** Management Software helps in maintaining and configuring the infrastructure.
- **Deployment Software:** Deployment software helps in deploying and integrating the application on the cloud**.**
- **Network:** Network is the key component of cloud infrastructure. It allows connecting the cloud services over the internet. It is also possible to deliver the network as a utility over the internet i.e. the consumer can customize network route and protocol.
- **Server:** Server helps to compute resource sharing and offer other services like resource allocation and deallocation, monitoring resources, security etc.
- **Storage:** Cloud uses distributed file system for storage purpose. If one of the storage resources get fail then it can be extracted from another one, which makes the cloud computing more reliable**.**

## 1.4 SERVICE MODELS

- **Cloud Computing is built on Service Models [9], which are reference models. As described below, these may be divided into three fundamental service models:**
- Infrastructure as a Service (IaaS)
- Platform as a Service (PaaS)
- Software as a Service (SaaS)

There are other more service models, all of which may be packaged as XaaS, or "Anything as a Service." Identity as a Service, Network as a Service, Business as a Service, Strategy as a Service, and Database as a Service are examples of this type of service.

Infrastructure as a Service (IaaS) is the most fundamental level of service. It service model (Iaas) makes use of the underlying service model (Iaas), which means that each inherits the security and management mechanisms from the underlying model, as illustrated in the diagram:

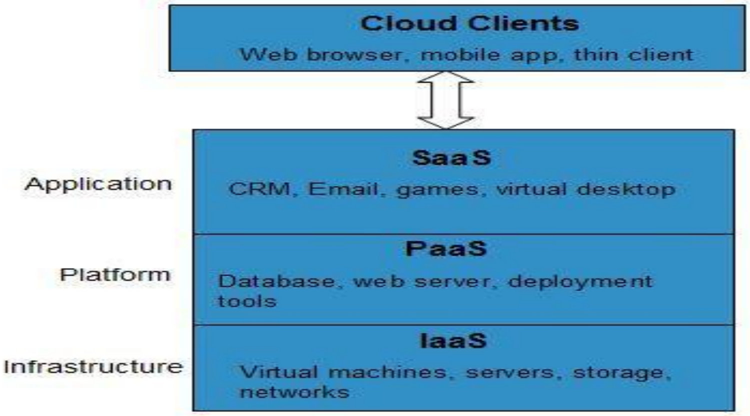

Cloud Clients
Web browser, mobile app, thin client
Application
SaaS
CRM, Email, games, virtual desktop
Platform
PaaS
Database, web server, deployment tools
Infrastructure
IaaS
Virtual machines, servers, storage, networks

**Fig: 1.4 Cloud Service Models**

### 1.4.1 INFRASTRUCTURE AS A SERVICE (IAAS)

IaaS gives users access to basic resources including real computers, virtual storage, and virtual machines. In addition to these resources, the IaaS provides:

Virtual machine disk storage

- Virtual local area network (VLANs)
- Load balancers
- IP addresses
- Software bundles

Server virtualization makes all of the aforementioned resources available to end users. Furthermore, clients have access to all of these resources as if they were their own.

#### 1.4.1.1 Benefits of IaaS Service Model

- IaaS allows the cloud provider to cost-effectively and flexibly locate infrastructure across the internet. The following are some of the advantages of IaaS.

  **Full Control of the computing resources by Administrative Access to VMs**

  IaaS allows the consumers to access computing resources by administrative access to virtual machines in the following manner:

  - The Consumer issues administrative command to the cloud provider to run the virtual machine or to save data on cloud's server.
  - The Consumer issues administrative command to virtual machines they owned, to start the web server or installing new applications.

**Provides Flexible and Efficient renting of Computer Hardware**

Virtual machines, bandwidth, IP addresses, storage, monitoring services, firewalls, and other IaaS resources are rented out to customers. The consumer is responsible for paying based on how long they keep a resource. The customer may also run any programme, even a bespoke operating system, thanks to administrative access to virtual machines.

**Provides Portability, Interoperability with Legacy Applications.**

It is possible to maintain legacy between applications and workloads between the IaaS clouds. For example, network applications such as email server, web server that normally runs on consumer-owned server hardware can also be run from VMs in IaaS cloud.

### 1.4.2 PLATFORM AS A SERVICE (PAAS)

PaaS provides the application runtime environment. It also includes the development and deployment tools needed to create applications [10]. PaaS includes point-and-click technologies that make it possible for non-developers to construct web apps.

• Google's App Engine and Force.com are two examples of PaaS providers. Developers may access these websites and construct web-based apps using the built-in API.

• Using PaaS has the downside of locking the developer into a single provider. An application developed in Python against Google's API and run on Google's App Engine, for example, is likely to work solely in that environment. As a result, the most serious issue with PaaS is vendor lock-in.

#### 1.4.2.1 Benefits of Paas Service Model

- **Lower Administrative Overhead:** There is no need for Consumer to bother much about the administration because it's the responsibility of cloud provider.
- **Lower Total Cost of Ownership:** There is no need for Consumer to purchase expensive hardware, servers, power and data storage.
- **Scalable Solutions:** It is very easy to scale automatically based on the application resource demands.
- **More Current System Software:** The cloud provider has the responsibility to maintain the software versions and patch installations.

### 1.4.3 SOFTWARE AS A SERVICE (SAAS)

The Software as a Service (SaaS) approach offers end users software applications as a service. The programme is housed on a server and is accessible over the internet in this case[11]. There are several SaaS apps available, some of which are listed below:

- Billing and Invoicing System
- Customer Relationship Management (CRM) applications

- Help Desk applications Human Resource (HR) Solutions

SaaS provides the Application Programming Interface (API) which allows the developer to develop a customized application.

### 1.4.3.1 Characteristics of SaaS Service Model

The characteristics of SaaS service model are:

- SaaS makes the software available on the internet.
- SaaS applications are cost-effective as they do not need any maintenance at the end user side.
- Software are available on demand.
- They can be scaled according to demand.
- They are upgraded and updated automatically.
- SaaS offers share data model. Therefore multiple users can share single instance of infrastructure. There is no need to hard code the functionality for individual users.

### 1.4.3.2 Benefits of SaaS Service Model

Using SaaS has proved to be beneficial in term of efficiency, scalability, performance and many more. Some benefits are given below:

- Modest Software Tools
- Efficient use of Software Licenses
- Centralized Management & Data
- Platfrom responsibilities managed by the provider

## 1.5 ADVANTAGES OF CLOUD COMPUTING

• **On-demand self-service** – Because there is no need for human interaction at the service provider, resources may be delivered to the consumer in a fully automated manner. The ultimate objective is to deliver a resource to a consumer "instantly," whenever they require it.

• **Broad network access** – Capabilities are available via the network and can be accessed by a wide range of end-user terminal devices, such as desktop thick clients and mobile thin clients (phones, laptops, tablets etc.).

• **Resource pooling** – The service provider builds the physical infrastructure so that all resources are grouped together in one or more pools. Customers are frequently unable to define the precise geographical location of the resources allotted to them from the pool. In fact, most providers give some high-level location options, such as a geographic region or a data centre (for example- USA vs. Europe, or USA west coast vs. USA east coast).

**Rapid flexibility** – Users should be able to quickly assign and release capacity when applications demand it. Elasticity is a two-way street, which means that applications must be able to both assign additional resources and release them when they are no longer required.

- **Measurable service –** Resources are priced at a finer granularity (months/years vs. hours/days) as they are utilised by customers, with suitable units for each resource (GB for memory, GHz for CPU, Mbps/Gbps or GB of transfer for network, and GB/TB for mass storage, for example).

- **Cost-cutting—** Businesses are frequently faced with the dilemma of increasing IT capability while reducing capital costs. By acquiring only the proper number of IT resources based on demand, an organisation may avoid purchasing superfluous equipment.

- **Application Programming Interfaces**— The software that enables machines to connect with cloud software in the same manner as the user interface permits interaction between people and computers is provided by the cloud computing system.[12]

## 1.6 DISADVANTAGES OF CLOUD COMPUTING

Although Cloud Computing is a great innovation in the world of computing but there also exist down sides, some of them are discussed below:

- **Security & Privacy:** The most serious fear about cloud computing is this. Because cloud infrastructure management and data management are provided by third parties, it is always risky to entrust sensitive data to them. Although cloud computing manufacturers claim that password-protected accounts are safer, any evidence of a security breach can result in the loss of clients and company.
- **Lock-in:** • One of the biggest drawbacks of cloud computing is the implied reliance on the service provider. This is referred described as "vendor lock-in," because switching from one supplier to another is difficult, if not impossible[13]. If a consumer wishes to move providers, the process of transferring large amounts of data from the old provider to the new provider can be extremely unpleasant and time-consuming. As a result, while selecting a vendor, one should carefully consider all of the possibilities. As a result, switching from one Cloud Service Provider to another is extremely difficult for clients. As a result, the service is reliant on a single cloud service provider.
- **Isolation Failure:** This risk involves the failure of the isolation mechanism that separates storage, memory, routing between the different consumers.
- **Insecure or Incomplete Data Deletion:** It is possible that the data requested for deletion may not get deleted. This happens because extra copies of data are stored but are not available.

**Growing Environmental issue of Global Warming:** Large-scale virtualized compute and data centres are becoming more widespread in the computing industry as a result of the introduction of Cloud computing. These distributed systems make use of mass-produced commodity server hardware, which is equivalent in theory to many of today's fastest supercomputers. However, merely running these systems consumes enough energy to power a city, and they require similarly large cooling systems to maintain the servers at normal operating temperatures [14]. This results in CO2 emissions, which adds considerably to the rising environmental problem of global warming.

## 2 Related Work

In 2022, Xing, *et al.* [1] The overall power consumption of physical machines (PMs) and switches, as well as the total network bandwidth resource consumption among VMs, are both reduced in this paper's formulation of a virtual machine placement (VMP) issue. We propose an energy- and traffic-aware ant colony optimization (ETA-ACO) approach to solve the problem. An energy- and bandwidth-aware PM selection system, a traffic-based VM ordering scheme, and a direct information exchange scheme are all developed to improve the performance of ETA-ACO. When choosing a PM to host a specific VM, the first approach has two phases. PMs with lower power consumption are preserved in the first stage. The one with the lowest bandwidth resource use is picked to host the VM in the second stage. In the second scheme, ETA-ACO ranks VMs according to their traffic demands in declining order.

In 2022, Kumar, *et al.* [2] Massive advancements in online, mobile, and computer technology, as well as their users, are developing at an exponential rate. Now is the age when the term "cloud computing" is very much in the public consciousness. Cloud services are gaining popularity among consumers (storage space, computational power and standalone applications). As a result, cloud service providers (CSP) are concerned about their clients' quality of service (QoS). Task scheduling was introduced to put it into effect. The primary purpose of task scheduling is to ensure that both the server and its clients achieve their objectives. Traditional task scheduling is insufficient to provide the optimum results. As a result, meta-heuristic techniques are needed to produce a solution that is close to optimal. This optimal solution determines the task-to-resource mapping and produces outcomes that meet the desired goals. Many researchers have been using meta-heuristic centric task scheduling algorithms such as ant colony optimization (ACO), particle swarm optimization (PSO), grey wolf optimization (GWO), whale optimization algorithm (WOA), and flower pollination algorithm (FPA) for developing new techniques in the last decade.

In 2022, Arora, *et al.* [3] Due to innovations such as on-demand processing, resource sharing, and pay-per-use, cloud computing is one of the fastest-growing topics in computer science. Security, quality of service (QoS) management, data centre energy use, and scale are all concerns with cloud computing. Scheduling is one of the most difficult challenges in cloud computing, since it requires several activities to be assigned to resources in order to improve service quality criteria. In cloud computing, scheduling is a well-known NP-hard issue. This will necessitate the use of an appropriate scheduling method. Several heuristics and meta-heuristics techniques were presented for optimally allocating the user's job to the cloud computing resources. In cloud computing, hybrid scheduling techniques have grown prominent. In this research, we looked at hybrid algorithms, which are algorithms that combine two or more algorithms and are utilised for cloud computing scheduling. The primary concept underlying algorithm hybridization is to extract relevant characteristics from previously utilised methods. In addition, this paper characterises hybrid algorithms and examines their goals, QoS parameters, and future perspectives for hybrid scheduling algorithms.

In 2022, Majumder, *et al.* [4] In the realm of computing, cloud computing is one of the newest technologies. It has demonstrated the viability of computing as a service. Individual consumers as well as large-scale commercial enterprises are progressively shifting to the cloud. The number of requests for cloud services is steadily growing. For appropriate service delivery, it is becoming increasingly vital to distribute incoming loads optimally across multiple virtual machines. In this paper, a dynamic load balancing scheme is proposed, in which group jobs are distributed according to priority to competent groups of virtual machines based on performance tier, and then queued jobs are distributed at the second layer in each group of virtual machines, taking into account network bandwidth, current load, memory capacity, and security measures.

In 2022, Goel, *et al.* [5] Data owners must upload their data to the cloud in the cloud computing paradigm. Due to the greatest distance between devices and the cloud, there is an issue with latency, bandwidth, and jitter. To address cloud issues, fog computing was brought to the network's edge. To improve the quality of service (QoS) characteristics during data transfer between Internet of Things (IoT) devices and fog nodes, resource and task scheduling is required. In a fog environment, several optimization and scheduling techniques were created. Despite this, the fog environment has issues with efficiency, latency, cost, calculation time, and overall execution time. Previously, NP-hard issues were solved using PSO (particle swarm optimization) or ACO (ant colony optimization) methodologies. Various optimization methods, such as Dolphin Partner optimization, Grey wolf, Moth-Flame, Firefly, crow, and others, are available for such optimization procedures. On the other hand, a solution to the problem was constructed using a priority queue and a round robin scheduling method. The implementation of PSO, ACO on the cloud, and Fog is contrasted in this article using the iFogSim toolbox. In fog computing, the findings of QoS parameters makespan and cost demonstrate an improvement in QoS over cloud computing.

In 2022, Arul Sindiya, *et al.* [6] Task scheduling challenges become more complex in a Cloud Computing context due to the dynamic and unpredictable nature of the environment. As a key need for establishing QoS, it specifies the necessity for efficient task scheduling design and execution. Cloud providers may make the most money by properly utilising their resources. The optimal scheduling algorithm examines the resources given by providers for executing the tasks, rather than the task list received from consumers. We present a Dynamic Group of Pair Scheduling and Optimization (DGPSO) method in this research. The suggested DGPSO improves the performance of AWSQP by employing VM pair implementation and a three-level partition-based priority scheme. Low, medium, and high are the three tiers of the priority system.

The VM pairing is done based on the workload size. Communication time, system capacity, memory size, and processor speed are among the VM's characteristics for this.

## 3 Need of Study

In cloud computing systems, load balancing is a significant difficulty. In every situation, a dispersed solution is necessary. Because it is not always practicable or cost effective to sustain one or more idle services at the same time as meeting demand. Because the cloud is a very complicated structure with components dispersed across a large region, jobs cannot be given to appropriate servers and clients individually for efficient load balancing. Static and dynamic load balancing methods are two types of load balancing algorithms. Static algorithms are best suited for situations that are homogenous and stable, and they can deliver excellent results in these conditions. However, they are frequently rigid and unable to adapt to dynamic changes in attribute values during execution. Dynamic algorithms are more adaptable, taking into account many sorts of system attributes both before and during runtime. Load balancing is the technique of redistributing load among processors in order to improve system performance.

### Optimization as a Solution to cloud load Balancing

Ant Colony Optimization (ACO) is a novel heuristic strategy for addressing difficult combinatorial optimization problems that was recently introduced. The pheromone trail laying and following behaviour of actual ants that utilise pheromones as a communication medium inspired ACO. ACO is based on the indirect communication of a colony of simple agents termed (artificial) ants, mediated by (artificial) pheromone trails, similar to the biological example. In ACO, the pheromone trails serve as dispersed numerical information that the ants utilise to probabilistically create solutions to the problem at hand, and that the ants alter during the algorithm's execution to reflect their search experience. This ACO algorithm attribute may be completely leveraged to address the cloud load balancing problem, as Ants in the ACO algorithm enable genuinely distributed and parallel processing, which is critical to the cloud balancing challenge.

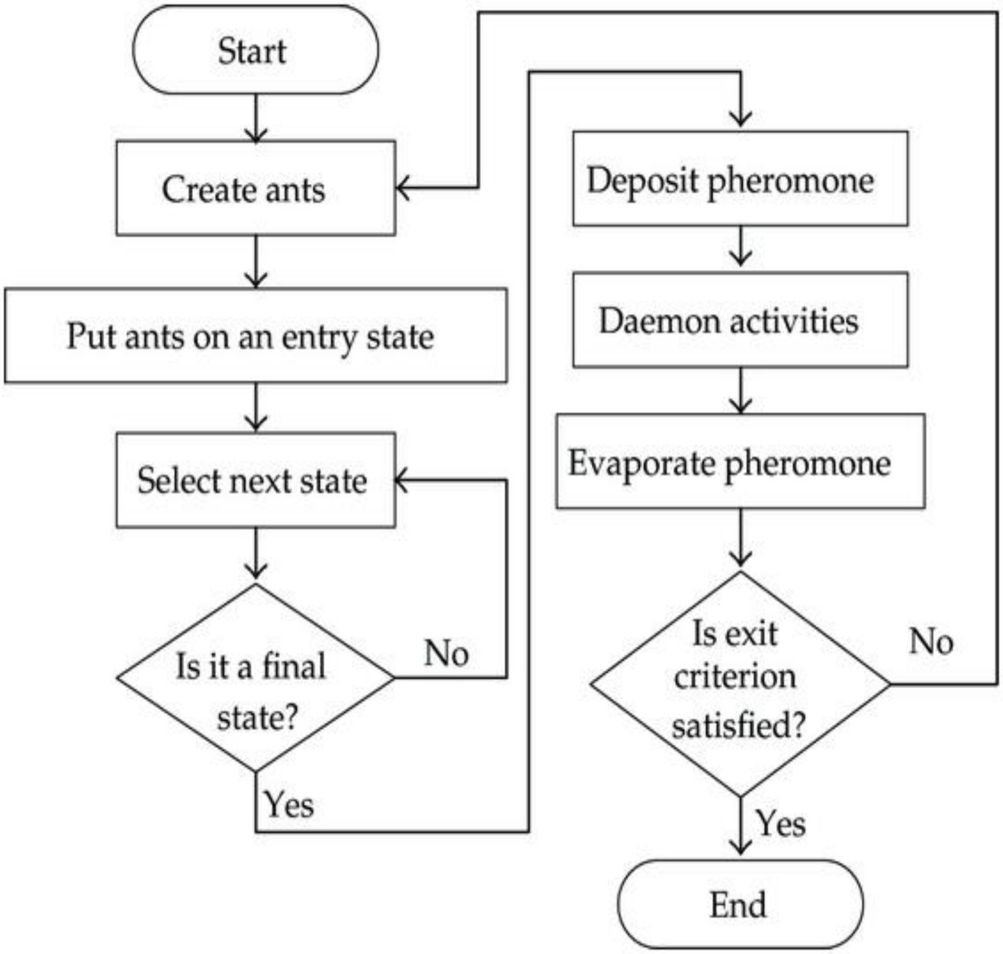
## 4 Research Goals/ Objectives

1 To characterize the Energy Model of a cloud system, as the Energy of a cloud system is defined to be the set of traffic loads under which the queues in the system can be stabilized.

2 To Study and Analyze the Ant Colony Optimization Algorithm and to evaluate the possibility of its applicability in Cloud based Systems

3 To implement the ACO algorithm in MATLAB and provide an Experiment to show that it is optimal for load balancing.

4 To Evaluate the Findings and Results via visualization Tools in MATLAB.



## 5 Conclusions

Until now, we've covered the core ideas of cloud computing and load balancing, as well as several current load balancing algorithms that can be applied to clouds. Furthermore, the solutions for closed-form minimum measurement and reporting time for single level tree networks with various load balancing schemes were investigated. Different techniques, such as the Genetic Algorithm and the Ant Colony, were compared. We have created a cloud system in this work, and the energy of a cloud system is specified to operate as a pair of traffic loads under which the queues in the machine are stabilised using ACO. We tested the expenses using the Ant Colony Optimization Algorithm and found it to be more efficient than other techniques like hereditary algorithms. We sincerely think that its implementation in Cloud-based approaches is viable, and that many options for extending load that is already in place to cope with ACO-based applications may be imagined. The load-balancer takes information about the Task from an individual and balances the cost accordingly. In the case when the load-balancer must somehow predict the likelihood of executing the optional parts, but requires additional communication, an ACO-based approach leads in less computationally demanding load-balancers. The cost, on the other hand, is quite limited.

## 6 Future Works

Load balancing is a job that is required in the Cloud Computing environment in order to maximise resource usage. We reviewed numerous load balancing strategies in this study, each with its own set of pros and cons. Because the application architecture in this research assumes a limited number of users, the load-balancing algorithm proposed maximises the percentage of optional material supplied. When considering a unique application, optimising the total number of requests served with optional material is another feasible aim that could be addressed in future research. The performance of cloud computing can also be improved if task dependencies are represented using ACO-based workflows.

## References

1. Xing, H., Zhu, J., Qu, R., Dai, P., Luo, S., & Iqbal, M. A. (2022). An ACO for energy-efficient and traffic-aware virtual machine placement in cloud computing. Swarm and Evolutionary Computation, 68, 101012.
2. Kumar, R., & Bhagwan, J. (2022). A Comparative Study of Meta-Heuristic-Based Task Scheduling in Cloud Computing. In Artificial Intelligence and Sustainable Computing (pp. 129-141). Springer, Singapore.
3. Arora, N., & Banyal, R. K. (2022). Hybrid scheduling algorithms in cloud computing: a review. International Journal of Electrical & Computer Engineering (2088-8708), 12(1).
4. Majumder, A. B., Majumder, S., Noor, D., & Das, P. (2022). A Two Layer Dynamic Load Balancing Algorithm Applied in Cloud Computing. In Applications of Networks, Sensors and Autonomous Systems Analytics (pp. 79-84). Springer, Singapore.
5. Arul Sindiya, J., & Pushpalakshmi, R. (2022). Job Scheduling in Cloud Computing Based on DGPSO. In Computer Networks and Inventive Communication Technologies (pp. 33-45). Springer, Singapore.
6. Goel, G., Tiwari, R., Anand, A., & Kumar, S. (2022). Workflow Scheduling Using Optimization Algorithm in Fog Computing. In International Conference on Innovative Computing and Communications (pp. 379-390). Springer, Singapore.
7. Ranbhise, I. S., & Joshi, K. K. (2014). Simulation and analysis of cloud environment. Simulation, 2(4).
8. Gupta, H., Vahid Dastjerdi, A., Ghosh, S. K., & Buyya, R. (2017). iFogSim: A toolkit for modeling and simulation of resource management techniques in the Internet of Things, Edge and Fog computing environments. Software: Practice and Experience, 47(9), 1275-1296.
9. Jo, M., Maksymyuk, T., Strykhalyuk, B., & Cho, C. H. (2015). Device-to-device-based heterogeneous radio access network architecture for mobile cloud computing. IEEE Wireless Communications, 22(3), 50-58.

10. Devi, T., & Ganesan, R. (2015). Platform-as-a-Service (PaaS): model and security issues. TELKOMNIKA Indonesian Journal of Electrical Engineering, 15(1), 151-161.
11. Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. Journal of network and computer applications, 34(1), 1-11.
12. Zaharescu, E., & Zaharescu, G. A. (2012). Enhanced virtual e-learning environments using cloud computing architectures. International journal of computer science research and application, 2(1), 31-41.
13. Kratzke, N. (2014). Lightweight virtualization cluster how to overcome cloud vendor lock-in. Journal of Computer and Communications, 2(12), 1.
14. Pierson, J. M. (2015). Large-scale Distributed Systems and Energy Efficiency: A Holistic View. John Wiley & Sons.