



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Task Scheduling for Cloud Computing Using Improved Genetic Algorithm

Satya Shekar Varma P

Assistant Professor, Dept. of CSE, MGIT, Telangana, India

**ABSTRACT-** Job scheduling is a very significant aspect of cloud computing. Different kinds of services are provided to the consumers within desirable throughput. A lot of optimization techniques emerged to enhance efficiency and reduce expenses. This paper provides an approach to an online mode heuristic algorithm named as ‘Genetic Algorithm’ and its comparison with the batch mode heuristic algorithms. Cloudsim is used to perform the simulation. The results of the simulation show that the Improved Genetic algorithm performs well compared to First Come First Serve Algorithm, Round Robin Algorithm, Shortest Job First algorithm and Genetic Algorithm.

**KEYWORDS:** Cloud Computing, Genetic Algorithms, Data Centre, Virtual machine, Cloudsim, Cloud Analyst, iCanCloud, GroudSim, NetworkSim, SPECI, and Cloud Report.

## I. INTRODUCTION

### Cloud computing:

Cloud Computing is the on-demand delivery of computing resources such as servers, storage, databases, software, networking, analytics, and intelligence over the internet with a pay-per-use model. The users are charged based on their usage hence no additional payment for unused infrastructure. Various virtualization techniques are used for real-time information exchange. There are a few most common Cloud deployments –Public Cloud - In Public cloud, the services are owned and operated by third-party service providers. These services are shared by multiple organizations. Private Cloud – Private cloud refers to resources used by a single organization or business. Services such as servers, networks, and data centres are completely dedicated to one organization. Hybrid Cloud – This type of cloud combines the public and private cloud and allows the data to be shared among them. There are two approaches an organization can follow here – Use the private cloud for some services and public cloud for others. Use the public cloud as a backup of the private cloud It provides three types of service models –Infrastructure as a Services (IaaS) - It provides services that include servers, networking, storage. It gives the highest extensibility and control over the management of computing resources. The flexible, innovative services are available on demand. It is similar to existing IT resources.

Platform as a Services (PaaS) - It is a service to the user that gives a platform to develop or test the application on the cloud. It removes the need for organizations to manage the underlying infrastructure. It reduces the complexity of middleware as a service.

Software as Services (SaaS) - It is a method of delivering software applications over the internet on-demand and on a subscription basis. In this model, the user has not to think about how the service is maintained or how the underlying infrastructure is managed. The user should think only about how to use particular software. Everything else is managed and maintained by the service provider.

The process of deciding which task will utilize the cpu time is called task scheduling. The scheduling of the task may be on the basis of their priorities. The priority assignment mechanism for the tasks can be either static or dynamic. In the case of static priority assignment the priority of task is assigned as soon as the task is created, thereafter the priorities cannot be changed. In dynamic assigning the priorities of the task can be changed during the runtime. The various scheduling algorithms followed are:

**1. First in first out (FIFO):** In this algorithm simply the task that came in first in the ready to run state will go out first into the running state. The task that goes out of the running state goes to the ready to run state. Figure shows the movement of the task in the ready to run and running state. Advantage – it is very easy to implement Disadvantage – we cannot have any priority mechanisms, in real time examples each task has a different priority and it is to be implemented, but using FIFO we cannot implement priority-based scheduling.

**2. Round robin scheduling:** In this case each task gets their turn after all the task are given their time slots. Thus it can be said that it is a time slicing wherein the time for each time slot is same and is given tasks one by one. The CPU time utilization for three tasks according to round robin is shown in figure. In this example, it is assumed that the time slots are 5 milliseconds each. The switching of tasks occurs in the following case: current task has completed its work before the completion of its time slot. Current task has no work to be done. Current task has completed the time slice allocated to it.

### 3. Shortest job first (SJF) scheduling:

In this case the task with the shortest execution time is executed first. This ensures a smaller number of tasks in the ready state. Thus, it can be said that the priority is higher for a task with lesser execution time and the priority of the task is lesser for the task with higher execution time. Disadvantage – if there were too many short execution time tasks, then the task with more execution time will never be executed. Advantage – the implementation of this scheduling method is also simpler as just the execution time of each of the tasks is ready to run state are to be compared to decide which task will be executed by the processor.

### Genetic Algorithm:

A genetic algorithm is a search heuristic algorithm inspired by Charles Darwin's theory of natural evolution. It reflects a process of natural selection where the fittest individuals are selected to produce offspring of the next generation. The process starts with the selection of fittest individuals from a population. They produce offspring which inherit the characteristics of parents. This offspring is added to the next generation. If parents have better fitness, then their offspring have a better chance of survival. This process iterates until a generation with the fittest individuals is discovered. We consider a set of solutions for a problem and select the set of best ones.

**Initialization:** The Genetic Algorithms require a population of  $n$  individuals to generate an optimal solution. Each individual is represented by a chromosome. From the  $n$  individual population, some individuals are selected based on some functions and operations are performed on them to find the fittest individual.

**Fitness Function:** This determines the productivity of individuals in the population. It represents the superiority and performance of one individual over the other. Low fitness value individuals are not selected for further operations. Fitness function is an important part of the Genetic Algorithm.

**Selection:** This is the process where the individuals are selected from the  $n$  individual's population based on a strategy which can be tournament selection, roulette wheel, selection based on rank. This is an intermediate solution for the next generation.

**Crossover:** In this operation, two chromosomes are selected and divided into parts then they are combined in such a way that the first part of the first chromosome is combined with the latter part of the second chromosome and vice versa. This improves the searching mechanism in the Genetic algorithm.

**Mutation:** This operation is used after the crossover step. If the crossover is not able to generate the required diversity due to crossover between the same populations repeatedly and hence mutation operation is done to bring some variation in the population. Usually, the value of mutation is kept very low. This operation alters the gene value in the chromosome. This helps the Genetic Algorithm to produce even more optimal results.

**Simulation:** Simulation is a demand for cloud computing to behold the implementation of real-time outlines. There are various tools available for it such as Cloudsim, Cloud Analyst, , GroudSim, NetworkSim, SPECI, and Cloud Report. We are using Cloudsim and Cloud Analyst in our project.

### 1.1 Problem Statement

Job scheduling is a very significant aspect of cloud computing. Different kinds of services are provided to the consumers within desirable throughput. A lot of optimization techniques emerged to enhance efficiency and reduce expenses. This application provides an approach to an online mode heuristic algorithm named as 'Genetic Algorithm' and its comparison with the batch mode heuristic algorithms. Cloudsim toolkit is used to show the cloud implementation and by enhancing the algorithms efficiency using Genetic Algorithms.

## 1.2 Proposed System

Our work offers a persuasive of the time-saving method. The online heuristic job scheduling algorithm called as a Genetic Algorithm is used here. We tried to balance the workload by arranging VM based on their processing power and arranging the cloudlets according to their Length The list of VM and cloudlets is then submitted to the broker for the allocation. Broker allocates through a Genetic Algorithm and allocation of resources is done.

Only two models have been used so a proper comparison cannot be give as an output performance so we have used FCFS, SJF, RR and Genetic Algorithm. The output is compared with all these algorithms and then the results are given out. The output for the genetic algorithm which has boosted the existing solution and produces more accurate results.

## II. RELATED WORK

**Task Scheduling in Cloud Infrastructure using Optimization Technique Genetic Algorithm, 2020:** The motto of this approach is to improve or satisfy the on-demand network access for all the customers from the pool of resources. Use latest genetic algorithms to solve all kinds of cloud-oriented systems like P2P, grid and many more and give a total output of it in the system. The proposed method achieves significantly better performance in comparison to traditional designed features. The complexity of this feature set is quite low, and could thus make the comparison unfair with the other models.

### **Review on Task Scheduling in Cloud Computing using parallel Genetic Algorithm,2020:**

The proposed method used a robust and easily adaptable search technique to minimize global execution and solve all cloud problems. The algorithm used in this research is Hybrid Genetic algorithms and more sophisticated ones to resolve that process. It has been used easily to for accessing the data and produce outcome. In proposed system, basic features of previous systems existed and also it solves problems in all domains. It cannot be compared to other as it is good in performance but also weak to handle errors.

## III. MODULE DESCRIPTION

### **CLOUDSIM:**

Cloudsim is a software which is used to perform modeling, simulation, and experiments in cloud computing services and infrastructures. "The tool is developed by Cloud Computing and Distributed Systems Laboratory at the University of Melbourne. We are using it to implement different job scheduling algorithms such as First Come First Serve algorithm, Round Robin algorithm, Shortest Job First algorithm, and Mainly Genetic algorithm. It can instantiate many data centers which consist of storage servers and physical host machines. These machines host multiple VMs executing several cloudlets. CloudSim can perform simulations of assigning and executing a workload on a cloud infrastructure.

### **CLOUDLET:**

Cloudlet in Cloudsim defined the workload, which is to be executed during the simulation run of the Cloudsim simulation engine. Cloudlet in Cloudsim is a model class that exists inside the package 'org.cloudbus.cloudsim'. Cloudlet is one of the most important models which defined the specifications for a simulation engine corresponding to the real-life candidate application to be considered for moving to a Cloud-based system. In contrast, we can also define this cloudlet as a single process or task being executed on the cloud-based system which is to be simulated through a Cloudsim simulation engine.

### **Cloud Broker or Datacenter Broker:**

This class model a broker, which is responsible for mediating negotiations between SaaS and Cloud providers and such negotiations are driven by QoS requirements. The broker class acts on behalf of applications. Its prime role is to query the CIS to discover suitable resources services and undertakes negotiations for the allocation of resources/services that can fulfil the application's QoS needs. This class must be extended for evaluating and testing custom brokering policies.



### 3.1 Design Methodology

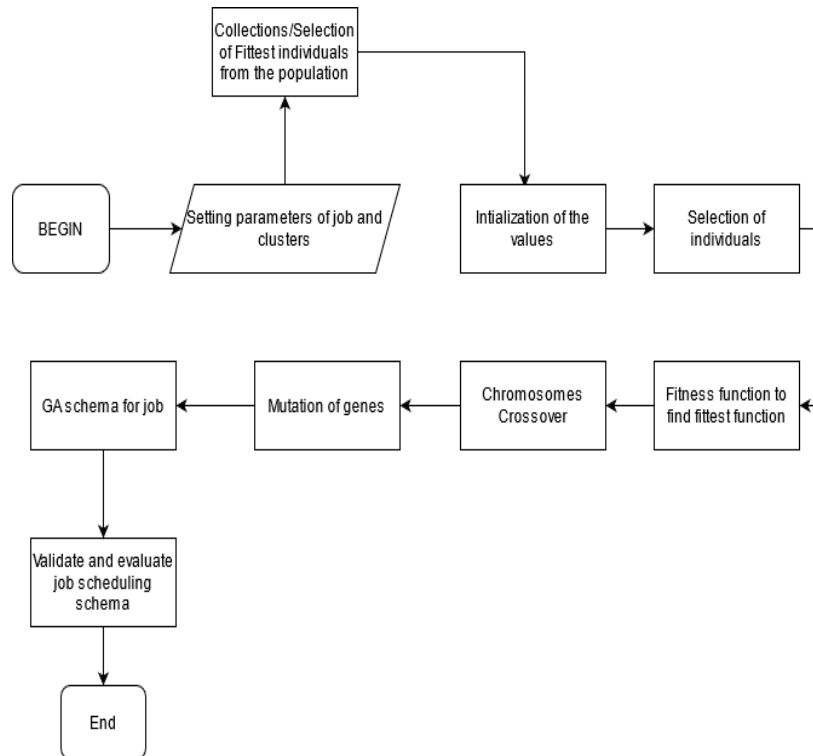


Figure: Block Diagram of genetic algorithm

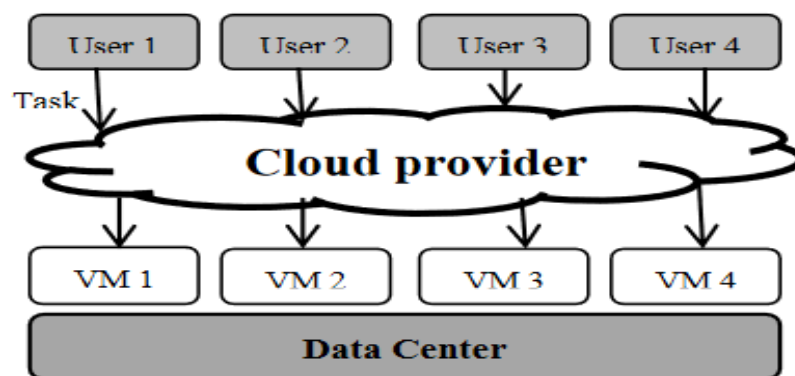


Figure: Communication between user and Cloud

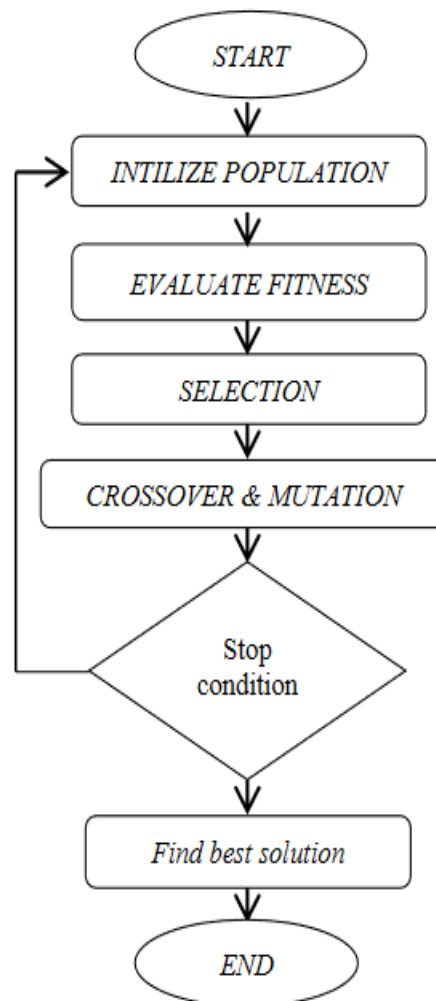


Figure: Genetic Algorithm Flow

### Description:

The process starts with the selection of fittest individuals from a population. They produce offspring which inherit the characteristics of parents. This offspring is added to the next generation. If parents have better fitness, then their offspring have a better chance of survival. This process iterates until a generation with the fittest individuals is discovered. We consider a set of solutions for a problem and select the set of best ones.

**Initialization** The Genetic Algorithms require a population of  $n$  individuals to generate an optimal solution. Each individual is represented by a chromosome. From the  $n$  individual population, some individuals are selected based on some functions and operations are performed on them to find the fittest individual.

According to the proposed TS-GA algorithm, the population is randomly generated using encoded binary (0,1). Therefore, the representation of solutions in task scheduling for each gene or (chromosome) consists of VM ID and ID for each task to be executed on these VM. Each VM and the executed tasks on it are encoded into the binary bit

### Fitness Function

The fitness function defines the relative importance of a design. A higher fitness value implies a better design. The fitness function may be defined in several different ways; it can be defined using the cost function value as follows:

$$F_i = 1 + \epsilon f_{\max} - f_i$$

where  $f_i$  is the cost function (penalty function value for a constrained problems) for the  $i$ th design,  $f_{\max}$  is the largest recorded cost (penalty) function value, and  $\epsilon$  is a small value (eg,  $2 \times 10^{-7}$ ) to prevent numerical difficulties when  $F_i$  becomes 0.

The main objective of the task scheduling is the cloud computing to reduce completion time for execution all tasks on the available resources. Therefore the task  $T_i$  on VM as  $VM_j$  as  $CT_{ij}$  is defined using equation

$$\text{Completion Time} = CT_{\max}[i, j]$$

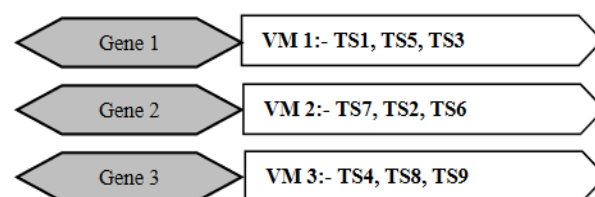
$$i \in T, i = 1, 2, 3, \dots, n$$

$$j \in VM, j = 1, 2, 3, \dots, m$$

Where  $CT_{\max}$  denotes maximum time for complete Task  $i$  on  $VM_j$ .  $n$  and  $m$  denote the number of tasks and virtual machines respectively. Therefore, to reduce the completion time which can be denoted as  $CT_{\max}$ , the execution time of each task for each virtual machine must be calculated for the scheduling purpose. If the processing speed of virtual machine is  $VM_j$  is  $PS_j$ , then the processing time for task  $P_i$  can be calculated by equation

$$P_{ij} = \frac{C_i}{PS_j} \quad (2)$$

Where,  $P_{ij}$  the processing time for task  $P_i$  on  $VM_j$  and  $C_i$  computational complexity of task  $P_i$



The processing time of each task in the virtual machine can be calculated by equation

$$P_j = \sum_{i=1}^n P_{ij}$$

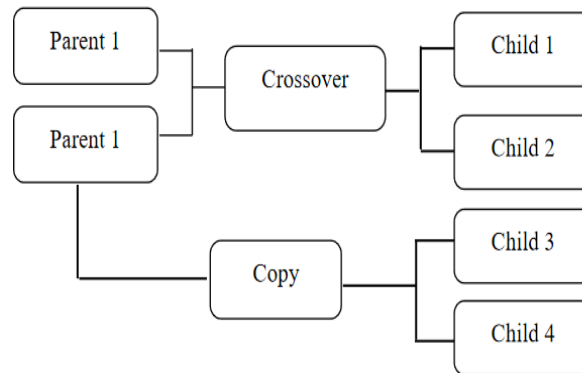
This determines the productivity of individuals in the population. It represents the superiority and performance of one individual over the other. Low fitness value individuals are not selected for further operations. Fitness function is an important part of the Genetic Algorithm.

### Selection

Selection is computationally more efficient and more amenable to parallel implementation. Therefore, the developed TS-GA algorithm, Tournament Selection is used to overcome the limitation of the population size. Two individuals are chosen at random from the population. A random number  $r$  is then chosen between 0 and 1. If  $r < k$  (where  $k$  is a parameter, for example, 0.75), the fitter of the two individuals is selected to be a parent; otherwise the less fit individual is selected. The non-chosen individuals are then returned to the original population and could be selected again

This is the process where the individuals are selected from the  $n$  individual's population based on a strategy which can be tournament selection, roulette wheel, selection based on rank. This is an intermediate solution for the next generation.

**Crossover** In this operation, two chromosomes are selected and divided into parts then they are combined in such a way that the first part of the first chromosome is combined with the latter part of the second chromosome and vice versa. This improves the searching mechanism in the Genetic algorithm.



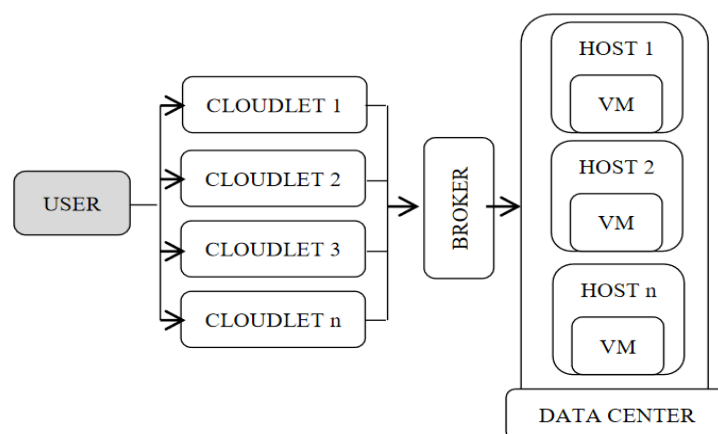
#### Keep Best Solution

There is a solution that might satisfy good fitness function, but it is not selected during the crossover process. By the proposed TS-GA algorithm, this solution is not removed from the population, but it is chosen and added to the population when next iteration is started. This step is considered as good step as some of the iterations can generate the best solution. Generally, according to the modified TS-GA algorithm, a set of modifications have been introduced.

These modifications areas follow

- The tournament is used instead of the roulette wheel in the selection process to select the best solution.
- The solutions not chosen in the selection process are considered and added to the new population. This might help in generating the best solution in the next generations
- The new crossover is introduced by considering parents individuals as new child (see Fig. 4)

**Mutation** This operation is used after the crossover step. If the crossover is not able to generate the required diversity due to crossover between the same populations repeatedly and hence mutation operation is done to bring some variation in the population. Usually, the value of mutation is kept very low. This operation alters the gene value in the chromosome. This helps the Genetic Algorithm to produce even more optimal results.



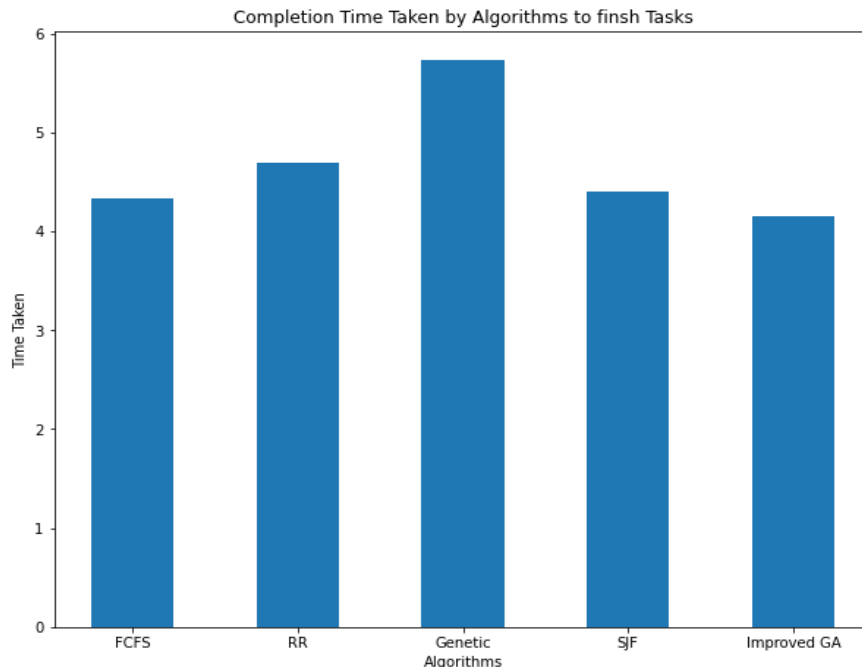
**Figure: The Cloud Sim Behaviour**

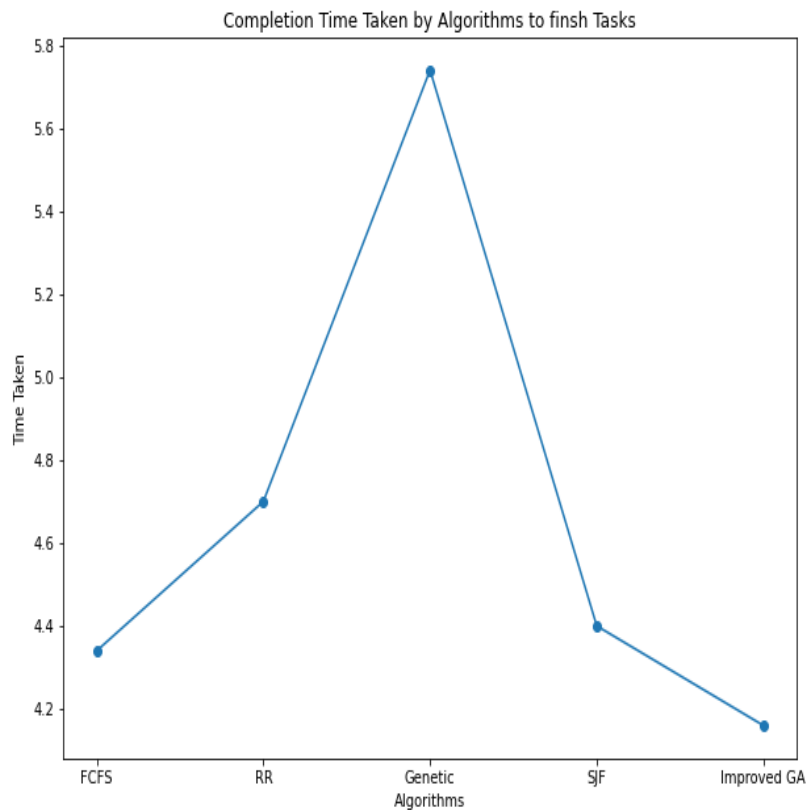


#### IV. RESULTS

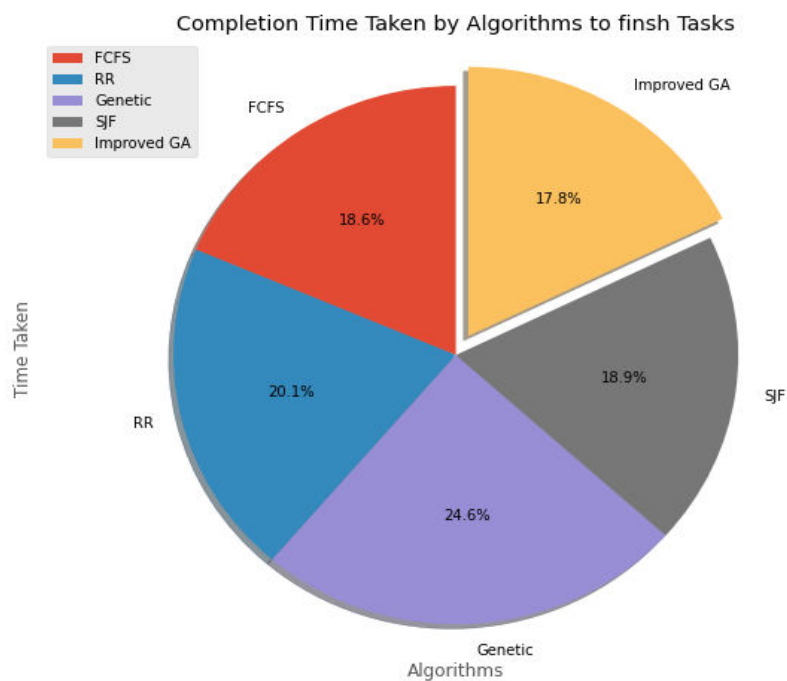
Here in the output we can observe that the cloudlet id signifies the available of unique station consisting the required resources the Status tell whether the process is successful. The Datacentre ID and VM ID signify the unique availability of the resources. The start time and end time tells about the time taken by the process to get the resource and finish the work. The FCFS takes a max time of 4.34 seconds to complete the work. Here in the output we can observe that the cloudlet id signifies the available of unique station consisting the required resources the Status tell whether the process is successful. The Datacentre ID and VM ID signify the unique availability of the resources. The start time and end time tells about the time taken by the process to get the resource and finish the work. The RR takes a max time of 4.7 seconds to complete the work. Here in the output we can observe that the cloudlet id signifies the available of unique station consisting the required resources the Status tell whether the process is successful. The Datacentre ID and VM ID signify the unique availability of the resources. The start time and end time tells about the time taken by the process to get the resource and finish the work. The SJF takes a max time of 4.4 seconds to complete the work. Here in the output we can observe that the cloudlet id signifies the available of unique station consisting the required resources the Status tell whether the process is successful. The Datacentre ID and VM ID signify the unique availability of the resources. The start time and end time tells about the time taken by the process to get the resource and finish the work. The FCFS takes a max time of 5.74 seconds to complete the work Here in the output we can observe that the cloudlet id signifies the available of unique station consisting the required resources the Status tell whether the process is successful. The Datacentre ID and VM ID signify the unique availability of the resources. The start time and end time tells about the time taken by the process to get the resource and finish the work. The IGA takes a max time of 4.16 seconds to complete the work.

#### Output Graphs:





Graph 4.2: Line Graph representation of Completion time taken by Algorithms to finish tasks



Graph 4.3: Pie Graph representation of Completion time taken by Algorithms to finish tasks.



## V. CONCLUSION

With the booming requests of cloud services today, there are issues of providing on-demand services and resources in a cloud environment and can't be overlooked. Our work offers a persuasive of the time-saving method. The online heuristic job scheduling algorithm called as a Genetic Algorithm is conducted as experimental studies.

We tried to balance the workload by arranging VM based on their processing power and arranging the cloudlets according to their Length The list of VM and cloudlets is then submitted to the broker for the allocation. Broker allocates through a Genetic Algorithm and allocation of resources is done.

In this paper, we did a comparative analysis of the results of job scheduling algorithms like First Come First Serve algorithm, Round Robin Algorithm, Shortest Job First Algorithm. The results of using the genetic algorithm are load balancing, improved processor utilization, make span minimization, cost minimization, and maximized throughput. The Genetic algorithm gives better results as compared to the batch heuristic algorithms.

## 5.2 FUTURE SCOPE

In the future we can improve this algorithms efficiency with more advanced genetic techniques. As we have used a cloud simulator we have to test it in real time cloud, so in the future we can try to do the task scheduling in real time cloud for a small scale of operations and try to modify the code to the appropriate results. In the future we can also take other algorithms in comparison to the genetic and get more accurate results and increase the performance.

## REFERENCES

- [1] C. K. Rath, P. Biswal and S. S. Suar, "Dynamic Task Scheduling with Load Balancing using Genetic Algorithm," 2018 International Conference on Information Technology (ICIT), Bhubaneswar, India, 2018, pp. 91-95.
- [2] M. Lagwal and N. Bhardwaj, "Load balancing in cloud computing using genetic algorithm," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 560-565.
- [3] Anshulika and L. A. Bewoor, "A genetic algorithm approach for solving a job shop scheduling problem," 2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, 2017, pp. 1-6.
- [4] H. A. Makasarwala and P. Hazari, "Using genetic algorithm for load balancing in cloud computing," 2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Ploiesti, 2016, pp. 1-6.
- [5] Genetic Algorithm chapter 2 and 3 shodhganga.
- [6] S. Yin, P. Ke and L. Tao, "An improved genetic algorithm for task scheduling in cloud computing," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 526-530.
- [7] F. Fakhfakh, H. H. Kacem and A. H. Kacem, "Simulation tools for cloud computing: A survey and comparative study," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017, pp. 221-226.
- [8] L. Zhao, Y. Dong and C. Huang, "A Study of Link Load Balancing Based on Improved Genetic Algorithm," 2013 Sixth International Symposium on Computational Intelligence and Design, Hangzhou, 2013, pp. 277-280.
- [9] Bhosale KK, Jadhav VD. A review of genetic algorithm for metal cutting processes and a research agenda. International Journal for Research in Applied Science & Engineering Technology (IJRASET). 2015;3(V).
- [10] Durairaj M, Kannan P. Improvised Genetic Approach for an Effective Resource Allocation in Cloud Infrastructure. International Journal of Computer Science and Information Technologies. 2015;6(4):4037-46.
- [11] S. S. Rajput and V. S. Kushwah, "A Genetic Based Improved Load Balanced Min-Min Task Scheduling Algorithm for Load Balancing in Cloud Computing," 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), Tehri, 2016, pp. 677-681.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.118**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details