CHAPTER 1
CHAPTER 1
INTRODUCTION

1.1 EVOLUTION OF COMPUTING

The computing field is one of the rapid growing industries that are fueled by the fast technological developments in the areas of hardware and software. The technological development in hardware includes chip development and fabrication technology, speedy and low cost microprocessors with high speed bandwidth connection and low latency interconnection networks. Recent improvement in Very Large Scale Integration (VLSI) technology is one of the important roles in the development of powerful sequential and parallel computers. On the other hand the software industry is also developing rapidly in terms of software like Operating System, programming languages, tools and development methodologies that are largely available in the market. This helps to develop various applications on scientific, engineering and commercial needs. The greatest challenge in this application is to identify the weather forecasting and earthquake analysis. For this purpose clients move towards the development of highly advanced parallel computers.

The computing era started with improvements in hardware designs, and then developed the system software applications. The computing system consists of three different phases like research, development, commercialization and commodity. The knowledge behind the advancement of computing system components is progressively matured in the field of parallel computing. It mainly focuses on technology advancement rather than maturity.

The important objectives of creating and using parallel system are to help the drawbacks faced by the speed of single processor. The companies produce parallel system to provide better performance and economy than sequential system. In the past era there were so many computing systems like scalable parallel computer architectures that gave high performance. The performance of the system depends on the processors, memory and interconnection. The growth of communication and
computing [103] industry begins from yearly 1960’s. In the year 1960’s mainframes computer were launched for the development of minicomputers. The Personal Computers (PCs) were evolved [13] during the year 1980’s. The workstations were launched in the year 1985 and utility as a computing was used in the year 2006.

**Cluster Computer Architecture**

Cluster computing is a kind of parallel or distributed processing system that contains group of interrelated stand-alone machines which are joined together as single machine. The following Figure 1.1 shows the architecture overview of cluster computer.

![Cluster Computer Architecture Diagram](image_url)

**Figure 1.1 Architecture of Cluster Computer**

Computer node has either single or multiprocessor system consists of I/O function, memory and an O/S. The inter connection of two or more computer nodes is called a cluster. If exist a single cabinet that can be separated and connected with the help of Local Area Network (LAN). LAN connected from a single computer node is used by clients, Cluster provides a cost-effective strategy that improves the salient features and improves the reliable and fast services. The only drawback is spending more expenses on sharing the memory. It consists of Multiple High Performance
Computers which include PCs, workstations and Symmetric Multi Processor (SMP). The state-of-the-art Operating Systems insist of Layered / Micro-kernel based on network that has High Performance Networks/Switches such as Gigabit Ethernet. Network Interface Card (NIC) has fast Communication Protocols and the Services like Active and Fast Messages. Cluster Middleware consists of Single System Image (SSI) and System Availability Infrastructure.

NIC is used to communicate, transmit and receive data packets between the cluster nodes with the help of switch. Among the cluster nodes the communication software provides speedy and reliable data communication between the nodes inside and outside. Cluster with special kind of network called Myrinet protocol is used for fast communication with its nodes. The Myrinet protocol potentially bypasses the operating system and removes the critical communication overheads by providing direct user-level access to the NIC. Cluster nodes consist of collection of computing resources which works on a single machine. The middleware is liable for offering an illusion of a single system image and accessibility consist of independent computers that are interconnected together. The programming environment offers portable, proficient and user friendly tools developed for applications. It consists of debuggers, profilers and message passing libraries.

**Clusters Classifications**

The cluster gives various characteristics like minimum cost, good performance, scalability, maximum throughput. It allows the organizations to improve their processing power with the help of using advanced hardware and installing and maintaining of software at reasonably low cost. The affordable upgrade path provides expandability of the organization increase their computing power, while preserving their existing investment and without incurring a lot of extra expenses. The performance of this application also improved with the support of scalable software environment. Another benefit of clustering allows a backup computer to take over the tasks of a failed computer located in its cluster.
Clusters are classified into various categories based on the following factors:

Application target the technological environment is based on High Performance (HP) clusters used in parallel computing. The main purpose of construct High Performance Application (HPA) cluster is to integrate the performance and availability of a single computer.

Node ownership is the difference between dedicated clusters and non-dedicated node based on the ownership in the cluster. In dedicated clusters, the individual does not possess any workstation. Dedicated cluster share the resources in a parallel computing manner that performs on the entire cluster, whereas in the non dedicated the individuals have their own workstation and applications are implemented using stealing idle Central Processing Unit (CPU) cycles. In fact, most cases the workstation CPU cycles are idle, at the peak hours. Adaptive parallel computing is dynamically altering the set of parallel computing in non-dedicated workstations. The major problem in non dedicated clusters, is sharing of application to run between the workstation owners and remote users. The client expects speedy interactive response from their workstation, while the latter is only concerned with speedy application turnaround by utilizing any spare CPU cycles. The workstation gives importance to sharing the processing resources erodes the concept of node ownership and introduces the need for complexities such as process migration and load balancing strategies. This kind of strategy allows clusters to provide sufficient interactive performance as well as to offer shared resources to demanding sequential and parallel applications.

The node hardware consists of PC, workstation or SMP.

- Clusters of PCs (CoPs) or Piles of PCs (PoPs)
- Clusters of Workstations (COWs)

1.2 CLOUD COMPUTING

Cloud computing, as a current profitable posing, started to become apparent in late 2007. It was planned to empower computing across widespread and diverse resources, rather than on local machines or at remote server farms. There are so many
definition for Cloud Computing, most of the authors suggested that it consists [58] of clusters of distributed computers called as Clouds that providing on-demand resources or services over a network with the scale and reliability of a data centre.

Cloud computing is a prospect technology [98] that need not compute on local host computers, it will operate on centralized facilities by third-party.

Grid computing serves as the basic for the establishment of cloud computing. In grid computing performs two types of services namely executing the jobs and allocating the resources. These types of services [24] are done with the help of scheduling algorithm as a basic tool. The better scope of the grid computing in computing technology is based on the scheduling.

Cloud computing is a computation approach [105] that deals with exciting flexibility, on-demand based service and virtualized resources. In Cloud computing the cloud users are connected by the service provider using internet. While the client asks any request [24] to service providers and they process the client request. It is a major role in cloud computing. The service provider takes low response time for processing the client is request with minimum throughput, this service is not guaranteed for all the users those are users of cloud.

It is one of the evolving knowledge that tops to the next generation of Internet. It is a web based distribution model for IT services that improves the offers enhanced and effective computing through enriched collaboration and availability, scalability and agility. The improved and competent computing is provided through a virtualized technology infrastructure, which is preserved and protected for the users. It offers subscription as pay-per-use basis and guarantee considerable reduction in the purchase of hardware and software cost. It also reduces the carbon footprints and minimizes the total energy costs.

Loganayagi, S.Sujatha [55] defines Cloud Computing as large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage,
platforms and services are delivered on demand to external customers over the Internet.

Software industry is not a static. Large amount of money is invested for purchase and maintenance of hardware and software. Without upgrading the system it increases the performance of hardware/ software requirement at cloud environment. Scalability and maintenance are easier in cloud computing.

It has been made a major change in the field of computing industry particularly in client service. The service provider is able to offer on-demand [24] phase based on dynamic resources. It is the newest approach of computing technology that creates many advantages to the clients like user centric service, Quality of Service (QoS) assurance, flexible and scalable for the on-demand phases.

One of the peculiar advantages of Cloud Computing is to reduce the software and hardware cost, need not upgrade the old version or purchase new one. Cloud Computing are more flexible and highly automated one in terms of re-allocation and migration. Sustainability and less amount of time taken for process are the other benefits of Cloud computing.

Cloud computing is a new practice in the Information Technology (IT) circle. It is a development of three computing combinations like parallel, distributed and grid computing. It delivers safe, speedy, expedient data storage and net computing service run by internet. Basically cloud computing distributes three kinds of services [77] like Infrastructure as a service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). These services are offered to users in a pay-per-use for the on-demand based category. It deeply concentrates on reliable service, error [94] free and scalable infrastructure for presenting internet created application services.

These services are made accessible on a subscription basis using pay-per-use model to customers, not bother about their geographic area. It is an economic based model for provisioning services and it simplifies [94] IT management way in an easy and most responsive based on the dynamic needs of the corporate.

36
Cloud computing comprised of very large number of virtual and physical servers, controlled by many hundreds of system is a huge task. Cloud deliver secure, speedy, convenient data storage and computing service cantered by internet. Cloud computing is the present commercial offering that provides computing across widespread and diverse resources on global machines instead of processing in local machine. It mainly focuses on deliver secure, reliable, fault free and scalable infrastructure for hosting Internet based application services. It is a cost saving economic model for provisioning services and it leads to IT management that can be made easier and more responsive to the changing needs of the business.

Buyya’s definition

"A Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and virtualised computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers”.

Layers of Cloud Computing Model

The Figure 1.2 given below shows the layered model [70] of cloud computing. It consists of various layers like a physical resource combined with core middleware is used for delivering information as a service to its clients. User level middleware layer is used for access the PaaS. The topmost layer concentrates on SaaS which assists the lower layer. The various services offer to 3rd party are IaaS, PaaS and SaaS.

User Level Middleware

It consists of web 2.0 interfaces focused on cost effective graphical user interface designed using internet. This layer used to create a platform for easy creation, deployment and implemented in cloud environment.
Core Middleware

It provides various services like quality system, negotiation, admission control, pricing, service level agreement management, monitoring, controlling, billing and accounting. Example for core middleware is Amazon EC2, Google Application Engine and Aneka.

System Level

The topmost layer in the system level consists of data center which consists of many hundreds of servers installed. Data center supplies the computing power to cloud computing architecture. This layer includes physical resources for storage and application servers. These servers [76] are controlled by higher level virtualization and allow them to share the resources among the servers. The Virtual Machine (VM) is responsible for fault tolerant and isolated security for cloud.
The various resources of Cloud Computing include the software, hardware components, legacy application and network resources. VM layer comprise of management layer, deployment layer and Xen supervisor. Management layer consists of Console, QoS, Negotiation, Pricing control, Service Level Agreement Management, Monitoring, Accounting Management and Billing. The fourth layer includes Programming tools, Computing, Database [55] for information management, queuing for scheduling, Storage the valuable data, Identity Services the client requirements, Map Reduce Service, Web 2.0 and Distributed Programming. The final layer is used to execute the client applications.

Cloud computing services are becoming integral part of people’s daily life. These services are supported by infrastructure known as Internet Data Center (IDC). It increases the number of clients and consumes more energy called as skyrocketing. Corporate and academic circles take more efforts to manage the energy of IDCs.

IDC is sustained infrastructure for cloud computing. It consists of many thousands of servers and IDC and it consumes huge Mega-Watts [42] of power for purpose of executing and cooling the system.

Cloud computing has developed into a progressively important study topic given the robust evolution and migration of many network services to the computational environment. Using huge amount of computing resources [17] is a difficult task. It is the most popular prototype in offering services over the Internet. It is a lively area of research, and the popularity of this prototype is increasingly fast. Companies like Amazon, IBM, Google, salesforce.com, Unisys and so on, offer cloud based services. The main advantage of cloud computing is capability to provide IT resources on-demand. It eliminates the over-provisioning and under-provisioning of resources which are commonly seen in the business that have extensively varies due to development / decline and seasonal peaks. The resources provided by them are storage, processing power of CPU and IT services. These kinds of resources are periodically and geographically varying distance based on their clients available.

A person who uses the cloud is called cloud user. Like small, medium or large scale organizations uses this kind of cloud service. A supplier or vendor who offers
service to their clients based on subscription for a fee is called cloud supplier / vendor.

Cloud agent or broker is the middleman or intermediary between the service provider and client. The agent is liable [1] for configuring the client system based on request of the clients for the purchase of resources.

1.3 TYPES OF CLOUD

The below Figure 1.3 depicts the three types of cloud computing [104] services offer to its clients based on their accessibility needs.

Public Cloud

Public Cloud is also known as External Cloud. It is the traditional system that offer, multi tenant infra based services to general public, here the resource allocation is dynamically provisioned with fine-grained utility computing and self-service based with the help Internet, through web applications / web services. For off-site 3rd party service, the resources are shared and payment system is based on their usage.

Private Cloud

Private Cloud is also known as Enterprise Cloud. It is the second type of cloud accessibility. The service provider in the recent years describes the offer provided by clouds that are emulated on private networks. This type of service has no drawbacks
because private cloud delivers the product in huge data security, corporate governance and perform failure free operation.

In the recent years huge number of small, medium and large business people get their resources from private cloud service provider. It makes profit oriented survival in the IT business or the service provider would be able to provide network at affordable price for private clouds at large scale.

In private cloud use, logical sense is given importance rather than the physical sense. An example for private cloud is PaaS offered by Microsoft's Azure Services Platform. It is unavailable in the on-premises deployment. In this prototype the company runs its own data center or infrastructure for the use of internal and partners of the company.

**Hybrid Cloud**

Hybrid cloud is a combination or mixture of both public and private cloud. It includes multiple internal and external cloud service provider implemented in most of the companies.

**Advantages of Public Cloud**

The various advantages of Public Cloud computing are on-demand access, infrastructure development, best resource usage, reasonable pricing and support 3rd party services. There is no need for any additional investment in the development of infrastructure. There is no additional purchase of hardware, software, power and setup etc. It provides service to the on-demand phase and reservation phase based on contract agreement between service provider and expectation of clients. The resources are utilised in an effective and efficient manner. Cloud is connected with the help of internet and the client shares their infrastructure from various time zones like 24X7. The service provider analyses the utility of cloud, based on usage increase the bandwidth, forecast the demand and supply. Finally it gives a quality service. The accessibility [13] of cloud service is also highly available to its 3rd parties.
Roles of Cloud Computing

There are various roles played by cloud computing as follows:

Cloud Service Provider

A cloud service provider is the owner of the business and controls all the activities regarding their service to its 3rd parties. The major drawback in traditional system is huge amount of capital investment for managing and payment is more expensive. The small business organization realizes the improved operational efficiency of servers and virtualization agility. Amazon.com is the first cloud computing service provider modernizing its data centers with huge computer networks. This innovative small process enters into the fast-moving groups add some additional features that are speedy and easy to operate at the time of accessing it. In the year 2002 Amazon Web Services launched its utility computing service.

Client

Clients are consumers who can access its cloud computing services. Cloud computing gives more importance to privacy, so the client has increased. The user has used this privilege of community effort to create a bill of rights.

Agent or Vendors

An agent or vendor who sells the products and services that helps to delivery, adoption and utilize of cloud computing.

Characteristics of Cloud Computing

Amount of capital invested in cloud computing is one of the important factors that determines the cost of access. The service provider takes steps to reduce the operational overheads. This kind of barrier, to enter into the cloud IaaS provided by a third-party, does not need to be procured for one-time or irregular intensive computing tasks. Payment based on usage of a fine-grained options reduces the need for IT skills during implementation.
Device and location independence that enable clients to access cloud computing using a web browser is not based on their geographic area. The client accesses the cloud service like infrastructure in off-site option from anywhere and those are provided by a third-party using Internet as a connecting medium.

Reliability of the service denotes the failure free operation of cloud that enhance through the use of multiple redundant web sites, it is suitable for organization continuity and disaster recovery. In many cases the cloud computing services are affected due to outages of IT companies and the corporate managers are little bit affected.

The real time system is self served and scalability is done with the help of dynamic based on-demand plans. The provisioning of resources on fine-grained does not affect the clients in peak time load. The performance of the system is monitored, stable and loosely-coupled design constructed using internet services as the system interface.

The main security is protecting the data. In every organization data is treated as like assets of the company. The security is improved due to security oriented focuses on resources that connected in centralization of data. It considers the raise in loss of control over certain sensitive information. In recent years cloud computing is more emphasis on security services. The service provider is able to devote resources to solve the security related issues that more number of clients cannot afford. The service providers maintain audit log that monitors the log access register to identify the complexity in accessing the data. Sustainability cannot be easily achieved. It attains with the help of effective resource usage. Improved effective system is needed to reduce the carbon footprint. Computers and its associated infrastructure are the major consumers of energy.

**Various Cloud Computing Platforms Components**

The below Figure 1.4 shows the cloud computing consists of various components [104] like clients, storage, application, IaaS, PaaS and SaaS.
Application

Cloud computing application influences the software design, often removing
the necessity to install. It runs the application on the client own system, thus relieving
the problem of software maintenance, ongoing task and support.

<table>
<thead>
<tr>
<th>IaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PaaS</td>
</tr>
<tr>
<td>SaaS</td>
</tr>
<tr>
<td>Storage</td>
</tr>
</tbody>
</table>

Figure 1.4 Various Cloud Computing Platforms components

IaaS

Infrastructure is one of the types of service offered by cloud. It delivers
platform virtualization environment service for its client. The two kinds of services
offered by IaaS are CPU Process and Storage of information. The infrastructure
based service is provided namely Amazon.com, Nirvanic, GoGrid and so on.

PaaS

Cloud computing provides PaaS that gives computing platform, solution in a
stack as a service, facilitates deployment of applications without any additional cost
and complexity of buying and managing the underlying hardware and software layers.
Illustration of PaaS is Google product like App Engine, Microsoft product like Azure,
Manjrasoft product like Aneka.

SaaS

The cloud service application consists of product, service and solution for the
problem that are delivered and used in real-time over the Internet". For an instance in
Web Services, "software system built to support interoperable system to system interaction with the help of internet", which may be accessed by another cloud computing components, software, e.g., Software plus service, or end users directly. Example for SaaS is SalesForce.Com.

Storage

Cloud service includes the delivery of data storage as a service. The clients are able to store the valuable data in the service provider site. The storage as a utility computing charged on per gigabyte per month.

1.4 CHALLENGES OF CLOUD COMPUTING

There are various challenges faced [103] by cloud computing as follows:

Security

Security is one of the key challenges in cloud computing. The essential features of security are protecting the valuable information. There is huge possibility of hacking the information by unauthorized users.

Workload

Workload distribution is another issue. Overload and under-load lead to poor resource utilization. Maintenance of optimum load balance is to avoid systems performance in idle.

Computational Economy

The main objective of cloud computing is to reduce the cost of access the service. Based on their usage, service provider collects the tariff. Pay per use is the cost mantra of cloud. Cloud Computing cannot predict accurately regarding the cost of access in fair and reasonable manner.
Resource allocation and distribution

The important function of scheduling is to allocate the resources properly to reduce the fragmentation. Every job and every client is important one. Here the idle system is converted into a busy system environment.

Uniform access

The clients would be able to access the network uniformly. Proper allocation of bandwidth distribution during the time of peak load hour is crucial. The users facing any problem leads to customer dissatisfaction. The user would switch over from current service provider to another one.

Network Management

The service provider frequently watches the present scenario of cloud computing. Congestion occurs to service provider who manages and updates day by day what the client expects and what the competitor is doing.

1.5 SCHEDULING

Scheduling is the critical process that decides the time and cost of access the service. The best potency of cloud computing is to compute the subscription cost based on practice as pay per use in terms of per hour or memory space based. Clients urge them a well-organized scheduling policy needed for processing the jobs. The main purpose of scheduling is to develop the system performance in terms of minimize the turnaround time and load stability of the system. If deadlock occurs during the time of computation, the entire process slows down i.e. balance the peak time and non peak time. Every resource has some set of cost and performance. Grouping method helps to optimize the jobs related to them. In parallel system, scheduling of jobs is crucial task. Scheduling strategy primarily focuses on increasing maximum profit rather than quality service. The factor influences the successful scheduling denoted with response time, throughput and usage of memory space.
The most important objective job scheduling is to pick the finest suitable resources in the Cloud for Cloud Computing User’s (CCU’s) jobs, by taking both static and dynamic metrics restrictions of CCU’s jobs into consideration. The main reason for effective job scheduling strategy is to assist improving the response time, throughput, utilization, mean response time, mean reaction time and mean slow down are the significant consideration in parallel systems. Effective job scheduler would be able to supply appropriate jobs to the processor for the reason of attaining quality system. The largest challenge is the completion of the competent quality system in Cloud Computing. Performance is one of the key factors in parallel job scheduling policy. Scheduling is an essential element in jobs allocation and distribution. It’s like a manager who assigns the work and controls them. Scheduling classified into two types namely local scheduling and global scheduling. In local scheduling, it determines how the processes resident on a single CPU is allocated and implemented, where as in Global / Meta scheduling policy allocates the resources to the multiple processors instead of single processor or to optimize a system-wide performance objective.

The motto of this proposed research is to schedule the jobs in an effective and in efficient manner that distribute the jobs over the cloud environment. QoS is not a single step process, it is stepwise refinement. It is difficult to specify the customer requirements. The proposed scheduling algorithm helps to remove the drawbacks in the existing scheduling algorithm. In the proposed job scheduling algorithm, the jobs are grouped together and improved by the Combinational Backfill Algorithm (CBA) using TQS technique. The important scope of the proposed research is to provide improved utilization of resources and enhanced performance application capable of offering various cloud based services. In addition to the basic services like SaaS, PaaS, IaaS, other services such as Storage as a Service and Desktop as a Service also come under the category of cloud computing services.

In Dynamic scheduling policy of cloud computing, the resources are allocated based on the on-demand and reservation category. In Ant Colony Optimization (ACO) the load-adaptive model is responsible for job allocation and scheduling. Dynamic job scheduling policy is one of the important issues solved in a lot of technical fields. Xin Lu, Zilong Gu et al., [105] proposed a load-adaptive time slot
Priority Queue algorithm to resolve the congestion of the real-time traffic. Scheduled the task groups in cloud computing platform, each resource has some set of cost and performance. With the help of job grouping, communication of coarse-grained jobs and resources are optimally computed. Usually jobs are scheduled by client requirements. Client demands a new scheduling policy to overcome the problems faced by network properties between client and resources. New scheduling strategies may use some of the easiest scheduling concepts to grouping them together with some network, give better solution that leads to more efficient job scheduling.

The scheduling policy consists of much number of steps. There are many strategies provided to improve the response time and throughput. The response time is the time taken by a service provider to respond its client for a particular service request. Cloud networking is a mixture, integration of cloud computing and virtual networking. The benefit of the solution provides cloud computing network become tighter integration of virtualization benefits at computing and networking levels. It provides connectivity service, advanced transport and networking mechanisms, with the help of internet.

1.6 METASCHEDULER ARCHITECTURE OF CLOUD COMPUTING

![Diagram](image_url)

**Figure 1.5 Internal Metascheduler Architecture**
The above Figure 1.5 shows the Internal Metascheduler Architecture [93]. The two kinds of scheduler are local scheduler and Meta / global scheduler. Local scheduler allocates the resources to the single processor. The main objective of Metascheduler is to maintain detailed information about workload of all the companies participated in the cloud environment. The user submitted jobs are entering into the cloud environment the Metascheduler takes the decision making regarding what job is to be executed.

Cloud computing includes multiple datacenter that has an enormous collection of scalable resources. Resources may be hardware, software, infrastructure, platform etc. Each datacenter consists of one or more VM. User jobs are mapped between those VM by metascheduler. Each VM has one or more Processing Elements (PEs) and jobs are scheduled to PEs by local scheduler.

![Overall Architecture of MetaScheduler](image)

*Figure 1.6 Overall Architecture of MetaScheduler*

The above Figure 1.6 shows the overall architecture of metascheduler [98] that has three entities namely data center, VM and PEs. In this architecture the cloud users can submit their request for job completion to the metascheduler. All tasks received by metascheduler and scheduling decision are made by querying the cloud
cluster about the availability of free nodes. Jobs are executed at the cloud cluster by local scheduler. Metascheduler is responsible for mapping the jobs between cloud users and cloud clusters.

All the requested jobs are scheduled by metascheduler based on the decision making. Client querying the cloud cluster about the accessibility of free nodes afterwards the mapped jobs are executed at the cloud cluster done with the help of local scheduler.

The advantage of Cloud Computing is to reduce the cost of hardware and software, need not upgrade or purchase new one. Cloud Computing is highly automated and flexible system in terms of relocation and migration. Sustainability benefit consumes less amount of energy. The important feature of Cloud Computing [8] billing is based on-demand of the user request called as pay per use. The Cloud service offers two kinds of schemes to its clients like on-demand based and reservation services. When compared the cost of on demand service is higher than the cost of reservation. Distributed processing and parallel processing are the basis for the development of Cloud computing used to store the data in to the cloud made up of various software and different protocols built for processing the information.

Cloud computing makes the business act as a repository consists of many layers and applies power, storage and various services. Cloud service provider primarily focuses on reliable and fault free system for the clients based on internet. In Goal oriented job scheduling, assign weight for current job based on waiting time. Mostly the user submitted jobs are scheduled based on the requirements. Existing scheduling policy has overcome by the new scheduling strategies that help to improve the performance. The proposed technique used to merge the network together to provide good result when compared to the traditional based scheduling.

Cloud computing has developed a vast improvement from the beginning to current. Cloud computing crossed a decade from beginning in conducting research [60] on virtualization, networking, SaaS, distributed and grid computing. It is a service oriented design that reduces the cost of access to gather the information of
the clients offer greater flexibility and demand based services and so on. The clients need not buy any additional hardware and software cost.

Cloud computing is a current distributed based computing model and it is attracted by customer due to save the energy with the help of central management of resources. This model occupies a major part in energy efficient mechanism. Cloud Computing is a rental based application that offers QoS in terms of user agreement. The simulator used for scheduling is done with the help of CloudSim for cloud computing. Cloud computing is one of the upcoming computing technologies that provides various types of application based on requirement specification of the clients. It is a business oriented centralized facility that controlled and operated by service provider, compute and storage purpose based on global machines. It improves the utilization of resources effectively in an economic manner when allocated. The major benefit obtainable by cloud is secure, flexible, scalable, QoS, cost effective and convenient one.

Cloud computing is one of the types of commercial offering that enables the service around the world rather than the run on local server. Many professionals said that cloud computing is a cluster consists of distributed computer with the help of internet. Cloud computing policy based on great amount of virtual and physical servers handling several hundreds of system is the most important concern. It is a cost saving prototype that delivers service to customer and makes IT management at ease and more responsive to the changing needs of the business.

**CloudSim Toolkit**

CloudSim is a simulator toolkit that helps in design and simulates the system applications for cloud computing. It helps in system design and behavioral prototype model that includes data center, VM and allocation of resources. It is a generic based application method extended with simplified model. It includes single and multi networked system. It acts a link between the scheduling policy and resource allocation strategy using VM based on web.
So many researches have been conducted so far in the Hewett Packard (HP) Lab in USA with the help of CloudSim. It is a well-organized resource allocation and fine management of data center in a prearranged way. CloudSim makes use of it by both the research and corporate belong to this environment. It is fine tuned in terms of performance enhancement. The important factor that decided by CloudSim is efficient time management in terms of least time taken to execute the application and more flexible in terms of heterogeneous applicability in Amazon EC2, Amazon EC3 and Azure from Microsoft with slight changes in programming based on requirements.

The various characteristics of CloudSim as follows:

(i) Supported large scale computing that performed in a single or connection of a multi node system.

(ii) It acts as a platform for design, service, distribution of resources and allocation policy.

(iii) It supports the prototype with the help of internet.

(iv) In the federated (multi process) architecture that connects both public and private cloud.

**Salient feature of CloudSim**

With the help of virtualization engine CloudSim develops and controls a single and multiple virtualized services in data center. Virtualization engine is used to switch over the time sharing / First Come First Serve (FCFS) to Space shared / Round Robin (RR) and vice versa for processing the jobs.

The author Rajkumar Buyya, Rajiv Ranjan [70] develop a unique architecture for cloud based service, measure its performance in terms of testing to identify its effectiveness. Boston University Representative Internet Topology Generator (BRITE) Topology is used for linking bandwidth and latencies in end to end
connectivity. The findings identified in large scale computing surrounding with little or no overhead in memory usage. It is simple, flexible or non-federated of one for resource provisioning.

**Basic Elements of CloudSim**

**Cloudlet**

It is an application based service that has pre-assigned instruction length, data transfer and life cycle. Application services like social network, content delivery and scheduling.

**Cloudlet Scheduler**

It is a share and processing power among cloud let in the VM. The two kinds scheduler namely space shared and time shared.

**Data Center / Cloud Broker**

The data center act as mediating communication between the SaaS and service provider. It is amazing on behalf of VM persist of obtainable memory, processor and storage volume.

**CloudSim Framework**

CloudSim is responsible for the control queue and step by step implementation of simulation events. Each job is generated by the CloudSim entity that has a runtime stored in the queue called as future event. Each job has a set of time to insert into the queue. Next the jobs are scheduled on each step removed from and transferred to deferred event queue.
Cloud Information service

It supplies resource register, pointing and finds out capabilities. Peer-to-Peer Cloud computing is one of the growing next developmental platforms for resolving various complex problems in the field of science, engineering and management related discipline. It is predicted to solve various types of scattering in across different types of organizations. Resource scheduling is a complex part in large-scale distributed systems and therefore tools like demands used for analyzing and fine-tuning the algorithms before applying them to the real systems.

![Diagram](image.png)

**Figure 1.7 Layered structure of CloudSim**

The above Figure 1.7 shows the layered structure of CloudSim [70] structural design and skeleton elements. SimJava is the lower layer of the CloudSim. It has separate event imitation engine that trappings the innermost part functionalities needed for higher-level simulation frameworks like queuing and processing of events, configuration of system basics like data center, broker, host and VM. It is employed
for linkage between the elements that communicate and hold simulation clock. The subsequently layer is GridSim that grasp up high level software equipment for modeling numerous Grid infrastructures such as networks and linked traffic profiles and the principal Grid components like resources, data sets, workload traces and information services.

The subsequent layer is supported by the CloudSim that has programmatically extended the central part functionalities showing by the GridSim layer. CloudSim preserves a model and replicates of virtualized Cloud based data center environs such as excited management interfaces for VM, memory supervision, storage space and bandwidth continuance. This level administrator has instantiation and execution of core entity like VM, hosts and data axis at the point of simulation time. This level is capable for instantiating and clearly managing a vast scale; these infrastructures comprise many hundreds of system devices. The most important trouble in this layer is provisioning of hosts to VM based on client wants, administer and handle the application carrying out and vibrant inspection. The cloud service supplier inspects the ability of wide range of policy in assigning the hosts that needed relay at this level by programmatically expands the core VM provisioning utility. It visibly shows the distinction at this level on how the host is dispersed to vary competing VM in the Cloud. Cloud host being part of collective among a numeral of VM that executes function based on customer defined Quality System requirements.

The User Code is the topmost level in the simulation stack that exposes configuration interrelated functionalities for hosts that has number of nodes and their necessity and so on. Applications like number of jobs required by VM with their purpose, types and mediator scheduling policy. A Cloud application developer generates three types of function like a mix of client demand allocations, application configurations and Cloud accessibility scenarios at this layer and executes vigorous tests based on the practice configurations previously supported within the CloudSim. Cloud computing is quickly developing research area, there is a deficient in the defined standard, tools and methods that can competently undertake the infrastructure and application level complexity.
The next prospect would be a number of research efforts both in academic circles and business towards essential core algorithms, policies and application benchmarking standard based on execution contexts. By extend the fundamental functionalities previously showing by CloudSim, researchers would be capable to carry out tests based on specific scenarios and configurations. Hence researches allow the growth of finest practices in all the decisive aspects related to Cloud Computing. The decisions made in CloudSim was being developed and widely used again in the already existed simulation libraries and frameworks. CloudSim decided to take gain of previously implemented and proven libraries such as GridSim and SimJava are the grip low-level requirements of the system. For illustration, by using SimJava while passing message and incident handling in SimJava, it avoids reimplementation. This saved us time and cost of software engineering and testing. Similarly the use of the GridSim framework permissible is to use again its implementation of networking, information services, files, users, and resources.

SimJava and GridSim have been widely utilized in conducting acerbic edge research in resource management by numerous researchers. In this simulation previously detected errors are fixed through negotiation. By reusing these long validated frameworks, SimJava are capable to focal point on decisive aspects of the system that are related to Cloud computing. At the same time captivating advantage of the stead fastness of components that are not straightly related to Clouds.

Simulation appears to be the only optimum way to gather algorithms on huge-scale distributed systems for heterogeneous available resources. The real system simulation performs well in functioning without making any employment of these system methods to avoid complex, by avoiding the overhead of co-ordination of real resources. The effective working of simulation in very large hypothetical problems that would otherwise, it requires involvement of a huge number of active clients and resources, which is very complex to interact and construct a large-scale research surroundings for analyzing purpose.

The main motto of work proposed by the author Rajkumar Buyya, Rajiv Ranjan [70] is to analyze the effective resource allocation techniques based on cost efficient through simulation. The simulator consists of numerous thousands of
resources and many lakhs of customer that need numerous requirements and investigate scalability of system. It has step by step procedure, effective resource distribution plan and satisfaction to the customers. The service provider interested in gathering information on local economy and the global positioning based on the time zone of a particular resource plays role in securing jobs under various pricing and demand or supply situations.

As such a huge-scale simulation clients are done in huge amount of computing power used by parallel and cluster computing systems. The simulation model was applied in the areas of biotechnology, astrophysics, network design and high-energy physics in order to study usefulness of our resource allocation techniques. The important impact is on the way resource allocation performed for solving problems on cluster and cloud computing systems.

CloudSim produces resource management and scheduling algorithms in an economic manner, that are keenly interested to work with this CloudSim project, that can advantage directly in modeling efficient Electronic Commerce (E-Commerce) paradigm for service oriented computing. One of the potential outcomes of this project would be a programmable software toolkit that can be used by the researchers to analyze policies on large-scale distributed computing systems including the Internet, E-Commerce, Electronic Trading (E-Trading) etc.

**Main Features & Functionalities of CloudSim**

The CloudSim toolkit helps in designing simulation of entities in both the parallel and distributed computing. It is used to modeling and appraises the scheduling algorithm by the clients in terms of applications, resources, and resource agents. It helps in a complete ability for generate various classes of different types of resources that can be combined by means of resource agent, for resolving calculation and data intensive applications.

The resource can be a solo processor or multi-processor with common or dispersed memory and administer by time or space shared planners. The handing out
system within a resource can be mixed in conditions of dealing out ability, design and ease of use.

**Cloud Modeling and Simulation Kit**

The central part of the hardware infrastructure services is connected to the Cloud in the form of simulator by a Datacenter element for managing service requirements. These requirements are application components that are sandboxed inside the VMs, which need to be owed as a share of handing out power on Datacenter’s host elements. In VM represent a set of operations interrelated to VM life series of provision in the host for VM creation, destruction and migration. A Datacenter is collected by a set of hosts, which are in charge for administration VM at some point in their life series. Host is a component that stands for a physical computing node in the Cloud. VM dispenses a pre-configured processing capability expressed in terms of Millions of Instructions Per Second (MIPS) for memory management, storage volume and scheduling strategy.

The host element implemented in the interface that carries prototype of simulation done in cooperation of single core and multi core nodes. Distribution of application VM to Hosts in Cloud base data center is the accountability of the VMs Provisionary element. This element exposes a quantity of custom technique for researchers, which assists in accomplishment for fresh VM provisioning rule based on optimization purpose based on user or system centric. In FCFS base rule is employed by the VM Provision forwards the plan that distributes to Host. The system constraint such as the necessary amount of processing cores, memory space and storage allocation are requested by the Cloud consumer. For this purpose the concept of mapping is used.

Further complex procedure can be printed by the researchers stands on the infrastructure and application demands. In favor of each Host element, the allotment of processing cores to VM is completed based on the host allotment. The policies obtain into account how several processing cores are to be assigned to each VM and how a large amount of the processing cores can be able to efficiently be credited for a given VM. Thus, it is probable to allocate detailed CPU cores to exact VM in a space
shared strategy or to vigorously distribute the capability of a core amongst VM in
time shared plan and allocate cores to VM on demand or to identify other strategy.

Every Host element instantiates, a VM scheduler element that equips space
shared or time shared policy for allocate cores to VM. Cloud system designers and
researchers can expand the concept of VM scheduler element for conducting
experiment with some added custom allotment procedure. After that, the bigger level
information associated to the time shared and space shared procedure are depicted.

The CloudSim toolkit makes available a wide-ranging ability for simulation of
poles apart from classes of mixed resources, customer, applications, resource agent
and schedulers. It can be appropriate to simulate the application schedulers for solo or
multiple managerial domains dispersed computing systems such as Clusters and
Clouds. Application schedulers in the Cloud environment called resource brokers,
carry out resource discovery, selection and aggregation of a various set of distributed
resources for a single user. This denotes that every consumer has been in possession
of private resource agent and so it can be embattled to optimize for the necessities and
objectives of its vendor. In disparity schedulers supervision resources such as group
in a lone managerial field have whole control in excess of the rule making use of
allotment of resources. This way that every customer requires to put forward their
jobs to the central scheduler, which can be embattled to carry out overall optimization
such as advanced system consumption and generally customer satisfaction depending
on resource allotment procedure or optimize for high precedence user.

SimJava is a Universal purpose event simulation put together executing in
Java.Simulations. It includes a quantity of entities, each executes parallel in its
possession of thread. This significance assists in build a network of lively entities that
message by sending and receiving submissive event objects proficiently. The
chronological discrete incident simulation algorithm in SimJava is as pursue. Middle
entity Sim systems sustain a timestamp ordered wait in line of prospect actions. To
begin with, each and every entity is fashioned and their layer methods are set in
execute state. While a unit describes a simulation function, the Sim system entity
terminates the progress of that entity’s thread and seats an event on the expectations
waiting for processing the function. At the same time every unit has stop, Sim system
pops the subsequently result off the queue, moving forward the simulation time for that reason and starting again unit as suitable. This maintain until no additional procedures are created. But the Java Virtual Machine (JVM) supports native threads and then every unit initial at accurately the identical simulation time could run parallel.

**CloudSim Entities**

CloudSim entities ropes unit for simulation of lone processor and multi processor that have heterogeneous resources that can be configured as either with the help of time shared or space shared systems. It permits setting of the clock to poles apart time zones to reproduce geographic allocation of resources. It ropes entity that simulates association employ for message among resources. Throughout simulation, it generates an amount of multi threaded unit, a single processor that executes in parallel leads to a thread. A unit behavior requirement is to be simulated surrounded by its layer by SimJava. In simulation situation it desires the theoretical of all the entity and their time dependent relations in the real system. It requires carrying the creation of customer defined time dependent reply role for the work together entity. The reply utility can be a purpose of the precedent, present or both positions of entity. CloudSim base simulations hold entities for the consumers, agents, resources, in sequence service, figures and network based Input and Output (I/O). Initiation and execution of the request in CloudSim entity are discussed beneath.

**User**

All occasion of the User entity stand for a Cloud user. Every customer is different from the respite of users with respect to the subsequent individuality. Kinds of jobs to be produced are like the job execution time, numeral of metric replications, etc. the various scheduling optimization policy on the basis of reduction in cost, time or combination of both. Activity paces how it frequently generates fresh job, moment in time zone and completes the deadline and budget.
**Broker**

Every client is linked to an occurrence of the Broker / agent entity. A client submits the job to its agent and then the agent keeps an agenda for parametric jobs based on client scheduling plan every day. Earlier than scheduling the jobs, the agent vigorously gets a catalog of obtainable resources from the total directory unit. Each agent attempts to optimize the plan of its client and consequently, agent are predictable to features tremendous rivalry while in advance access to resources. The scheduling algorithms exercise by the agent must be greatly flexible to the marketplace situation based on demand and supply.

**Resource**

Single occurrences in the Resource entity signify a Cloud resource. Every one resource may perhaps be different from relax of the resources with respect to the following distinctiveness:

- numeral of processors;
- charge of processing;
- velocity of processing;
- inner process scheduling policy;
- Confined load factor.

The resource velocity and the job carrying out time can be definite in conditions of the ratings of standard yardstick such as MIPS and Specification (SPEC). In addition it is distinctive with respect to the regular mechanism. In the lead SPEC gains the resource contact particulars from the Cloud information service, agent can enquired resources straight for their stagnant and vibrant properties.

**Application Model**

Application model avoids several exact application design. It is awake of the service provider to describe it. The application design is used for designing the
parallel application such as process parallelism, Directed Acyclic Graphs method, divide and conquer methods etc. It can be well designed and simulated by means of CloudSim. It is an autonomous job may need changeable processing time and input files size. Such type of jobs can be produced and their needs are definite all the way through Cloudlet objects.

Cloudlet is a wrap up that includes all the information about the job and its carrying out manage particulars such as job length uttered in MIPS, disk I/O process, the size of I/O files and the task creator. These essential factors help in formative carrying out time, the time obligatory to carry input and output files among client and remote resources and persistent the processed Cloudlets reverse to the creator next to with the outcome. The CloudSim toolkit ropes an ample variety of Cloudlet management protocols and services that permit schedulers to chart a Cloudlet to a resource and supervises it during the life phase.

**Interaction Protocols Model**

The procedures for interface amid CloudSim entity are executing events that are used. These entities utilize the events of mutually service request and delivery. The executing events are increased by some entity to be distributed without delay or with particular delay to other entities or itself. Event begins from the similar entity called as internal events and those create from the external entities are called as external events. Entities can differentiate these events based on the source recognition related with them. The CloudSim procedure is used for defining entity services. Depending on the service procedure, the CloudSim events can be further classified into synchronous and asynchronous events. An event is said to be a synchronous while the event source entity remains until the event target entity carries out all the actions connected with the event. An event is said to be an asynchronous whilst the event source entity elevate an event and persist with other activities without waiting for its destination. Once the target entities take delivery of such events or service requests, it acts in response back with outcome by transferring one or more events, which can then acquire apt events.
It ought to be renowned that external events could be synchronous or asynchronous, but internal events need to be raised as asynchronous events only to avoid deadlocks. The CloudSim entity services like client, agent, resource, communication services and figures send from one event to other entities that suggest ask for service to distribute outcome or to increase internal events.

These services are exploiting to simulate a client with purpose; an agent for arrangement and an elective for report generator for building informational reports at the conclusion of a simulation. The event source and aim entities should have the same opinion upon the practice for service demand and delivery.

The procedure for interface among the clients defined and core entities are predefined. Once CloudSim initiates, the resource entities register themselves by sending events in the Global Information Service (GIS) and also processes the similar resource registration process GRIS (Global Resource Information Server) with the help of GIIS (Global Index Information Server) within the Globus system. Depending on the client entity’s demand, the agent entities propel an event to the GIS entity to indicate a question for resource detection.

The GIS entity precedes a catalog of list resources and their contact information. The agent entity propels events to resources with a demand for resource configuration and properties. GIS takes action with vibrant message such as resources price, potential, accessibility, stack and additional configuration factor.

These events connecting the GIS entity are synchronous in character. Depending on the resource choice and scheduling policy, the agent entity places asynchronous events for resource entities in regulate to transmit the Cloudlets for completing the agent’s requirement. The assigned Cloudlet does not wait for resources from GIS. When the Cloudlet processing is over, the resource entity renews the Cloudlet position and processing time and throws it reverse to the agent by increasing an event to show its finishing point.

The CloudSim resources bring into play internal events to simulate resource behavior and resource allotment. The entity requests to be model in such a move that
it is capable to accept each and every one event meant for it. Conversely, it is up to the entity to make a decision on the related actions. In favor of illustration, in time shared resource simulations internal events are programmed to suggest the end time of a Cloudlet, which has the least left over processing time necessity. Temporarily stipulation of external event arrives, they adjust themselves and share resource ease of use for each Cloudlet, which means the largest part recently scheduled event may not essentially denote the end of a Cloudlet.

1.7 OBJECTIVE OF THIS RESEARCH

The idea behind proposed research work is:

(i) To attain better consumption of resources and thereby achieve performance enhancements for an application capable of contributing various cloud based services.

(ii) To offer proficient QoS based on Dynamic Metascheduler for cloud computing.

This research work proposed to alleviate the problem of unbalanced allocation of jobs at the time of scheduling. These methods give better results than the existing algorithm.

1.8 SCOPE OF THE RESEARCH

The job scheduling policy that runs on parallel system using many kinds of metric to define the priority of the job and various scheduling limits. The key factor that affects the scheduling is more flexible, but determining its values is a difficult one. Rather than tuning low priority scheduling parameters it proposes the methodology grouping scheduling policies, which allow system administrators to specify only the high priority scheduling performance.

Recent job scheduling policies that execute on parallel computers employ many metrics for defining the job priority and various scheduling limits. Given that
several metrics may seem elastic, but determining their values is complex. In goal-oriented scheduling policies, that allow system administrator to a specific high-level scheduling only.

- Forwarding to get better the turnaround time and effectual use of all the jobs.

- Annoying to bring an optimized scheduling algorithm to attain the optimization for cloud scheduling problems.

In parallel system, the spotlight two scheduling algorithms based on performance prevent starvation and favors shorter jobs.

1.9 ORGANISATION OF THE THESIS

Chapter 1 presents the overview about the cloud computing, types, services, layered design and model of CloudSim. It deals with the quality system and efficient job allocation policy. It elaborates the history and evolution of job scheduling. It also gives basic idea about the problem.

Chapter 2 describes the brief overview of the existing state of the art in parallel system for designing a model and simulation. It introduces some background information of job scheduling and presents theoretic fundamentals of scheduling.

Chapter 3 depicts the Multi Queue Scheduling (MQS) method to improve the turnaround time and effective utilization of all the jobs. It exemplifies job allocation using MQS approach, which attempts to minimize the starvation during the allocation of jobs during the time of scheduling.

Chapter 4 explains the overall design of the newly proposed Optimized Resource Filling (ORF) approach to identify unused space in the scheduling and to achieve refilling the resource gap with the help of grouping approach.
Chapter 5 depicts the new approach called the Tri Queue job Scheduling (TQS) algorithm to avoid load imbalance during the distribution of jobs without blocking.

Chapter 6 describes the brief overview of the Reservation and On-demand Priority (ROP) based queue job scheduling algorithm.

Chapter 7 presents the overall design of Improved Priority based Queue Job Scheduling Algorithm for Cloud Environment (IPMQJS) algorithm that enhances the multi queue job scheduling. It explains the IPMQJS methodology to give priority for all the jobs in the queue based on their execution time jobs are classified into three types of priority namely high, medium and low.

Chapter 8 finally concludes with brief conclusive remarks and discussion on future research directions. It gives some observations and comments. The future enhancement and scope of the research are discussed briefly.

1.10 SUMMARY

An overview of Cloud Computing has been discussed in this chapter. This chapter highlights the objectives of maximum system usage and reduce fragmentation within the scheduling that increase the access speed and high resource usage which are to be considered for efficient job scheduling in cloud computing. The scope and the objectives of the thesis have been explained elaborately. Contributions of the thesis and organization of the thesis has also been presented. It presents the overview about the cloud computing, types, services, layered design and model of CloudSim. It deals with the quality system and efficient job allocation policy. It elaborates the history and evolution of job scheduling. It also gives basic idea about the problem.