CHAPTER 1

CLOUD COMPUTING: A REVIEW

1.1 OVERVIEW

Cloud computing paradigm [1] promises reliable delivery of services through data centres built on virtual computation and storage technologies and infrastructure. It will allow consumers to access data and applications, process their jobs using the "cloud" remotely, from anywhere around the world. In this chapter, firstly a brief overview of cloud computing covering its introduction, evolution, why cloud computing and cloud computing service models are elaborated subsequently followed by discussion on application and open challenges related to cloud and at last also detailing vendors of cloud computing are presented. Chapter after that presents performance indicator, research objectives, contributions and organization of the thesis.

1.2 INTRODUCTION TO CLOUD COMPUTING

Several computing researchers have attempted to define some of the components of computing like clusters, grids and clouds [2]. Attempts to define these components of computing basically target on some feature set, such as classes of technology, patterns of use, protocols etc. Here are some definitions that are detailed but generic enough to define these terms.

Clusters can be defined as a type of distributed and parallel system that is composed of a group of stand-alone computer systems, interconnected to work together as a single unit or a computing resource. On the other hand, Grids are a type of distributed and parallel systems that allows dynamic selection, sharing and aggregation of geographically distributed resources at runtime. Such dynamic selection of resources is made on the basis of their capability, cost, performance, availability and QoS (Quality of Service) requirements of the user [3] [4].

A cloud is a parallel and distributed system that consists of several virtual systems, interconnected in a dynamic manner to provide unified resources to a consumer depending upon the service-level agreement between the consumer and the service provider [5]. At a perfunctory glance, it seems like clouds are combination of the grids and clusters but it's far from truth. Clouds are, in fact, next-generation data centers that have virtualized nodes created via hypervisor technologies like virtual machines. These are dynamically provided on-demand and allow the end user to use it like a 'web service' for his/her requirements agreed upon by a service-level agreement [5].

The basic aim of cloud computing is to provide access of utilization of the immense power of distributed computation and storage. Such access of services would be on-demand. Consumers are promised 24-hour access to resources. As a result, the services need to be highly scalable, reliable, and dynamic with autonomic support. Consumers can indicate what level of service they require via QoS (Quality of Service) parameters, which are noted in the SLAs setup by the providers [1]. The Figure 1.1 represents cloud computing model and is shown under as

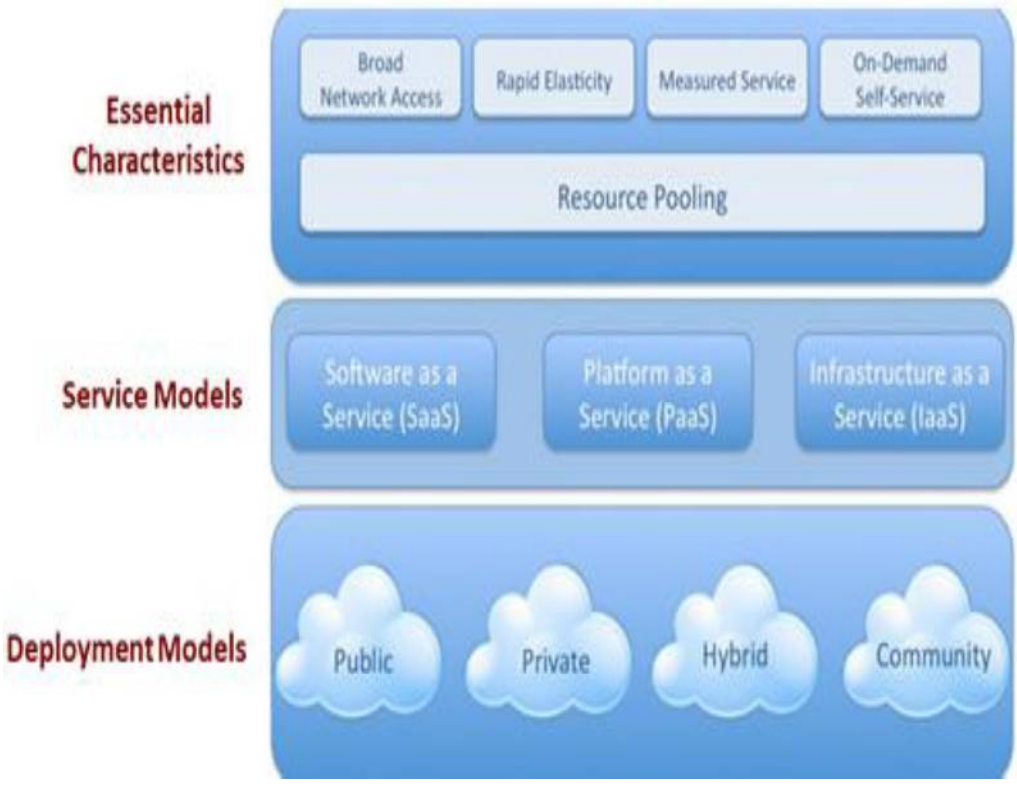

Figure 1.1 Cloud Computing Model [126]

Every service is loosely attached in cloud environment. If one service gets failed then it will not affect the other services. Cloud computing is performed in two phase as frontend and back end. The front end is a client who gets served by those services which are provided by the back end which is the cloud system [46] [82].

1.3 EVOLUTION OF CLOUD COMPUTING

It's the beauty of the advancement of society and technology that basic utilities are a phone call away from people. Basic utilities like water, gas, electricity and telephones are an inseparable part of our daily routines. Due to their frequent use and requirement, they must be readily available to the consumers. A consumer can use such utilities according to his/her requirements pay for availing such utilities according the respective usage. A chief scientist at ARPANET (Advanced Research Projects Agency Network) pointed out that computer networks are still in their early stages. As they grow, the concept of 'computer utilities' will spur up just like usual utilities in the modern world - telephones, electricity, cable etc. Going by this vision, it's expected that in the 21st century, computing industry will go through a massive change and the rise of computing utilities based on the on-demand service concept is inevitable. Just like the basic utilities, a consumer would be required to pay the provider on the basis of the usage. As a result, consumers will no longer need to invest in the expensive infrastructure to get access to certain computing services. There will be no need to setup, manage and maintain the IT infrastructure. Hence, software engineers are shifting their attention towards building software's that can be provided as an on-demand service instead of creating software's that will run only on individual computers. There are several computing paradigms that have been developed to meet this new demand of computing utilities, the top examples being grid computing, cluster computing and the very popular cloud computing [1].

1.4 WHY CLOUD COMPUTING?

Cloud Computing has been a favorite of several enterprises right from its time of inception. Cloud service providers have scaled up their businesses and are, now, providing several business solutions possible due to cloud computing. Enterprises are reciprocating by moving their IT infrastructure, services and applications to the cloud-

based architecture, which is helping these enterprises to improve their productivity and annual revenues. This signifies that the organizations must have an insight into the latest cloud computing technologies and the factors that may affect its growth [6]. The Figure 1.2 represents as why cloud computing should be used is shown below

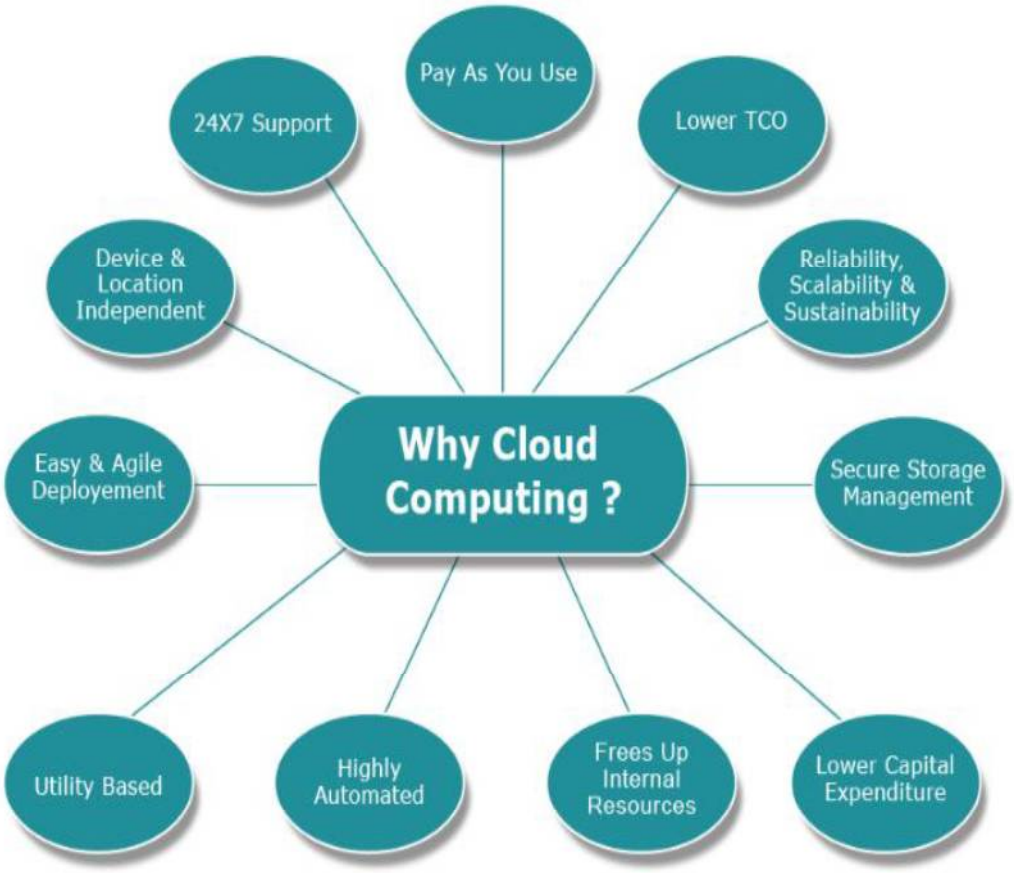

Figure 1.2 Why Cloud Computing is Used [126]

1.4.1 Economical and Technical Advantages of Cloud Computing

In business domain cloud computing occupying a major role in the business domain computing resources is delivered as a utility on demand to cloud clients over the Internet. The economic and technical advantages for cloud computing can be presented as under:

(a) **Rapid Elasticity**: Cloud computing is a flexible option as the user can request varied amount of resources (like storage size or no. of processors) depending on his requirements in this kind of environment. Such a feature makes it very easy

for web service providers to scale up or down the resources allotted to a user as they can add or remove the computing resources like servers depending on user demand [7] [8].

(b) **Pay-per-use Pricing Model**: Software systems are provided on the basis of pay only for what you use model [9] [10] [11]. As a result, it is innovative and flexible, allows round the clock access, simplifies the cost model and consumption model with faster access to applications and systems, secure, flexible, reduces maintenance cost and is resilient in the disaster recovery process [12].

(c) **On-Demand Services**: Several cloud services like server time, processing power, and network storage can be automatically configured and provided to the end user without requiring any human interaction with the service provider. This saves time and reduces hassles in scaling up or scaling down the use of such cloud services by the end user [7] [8].

(d) **Cost-Effective**: As cloud computing allows sharing of resources such as software, hardware, and IT infrastructure, the cost of the resources goes down. Such resources can be shared across multiple applications in same or different organizations, thereby offering organizations to bring a significant reduction in their capital expenses. Organizations no more need to invest in expensive infrastructure and can avoid repetitive expenses that happen due to the change in technologies over time [10] [11] [12] [14].

(e) **High Up-time**: Cloud computing offers high availability architecture that reduces or often, eliminates any downtime. It highly improves the services offered to the user [12]. The organizations can sign assured QoS (Quality of Service) with the service provider, according to the SLA (Service level agreement) [11].

(f) **Easy to Maintain**: As the service provider takes care of the software and infrastructure maintenance, organizations find cloud based resources very easy to maintain and it also brings down the IT manpower cost [12, 14].

(g) **Diversity of the Platform**: As the cloud infrastructure offers a variety of client operating systems and platforms, organizations can choose according to their requirement and aren't limited with one single platform[14].

(h) **Fast Rate of Development**: As the basic infrastructure is provided by the cloud service provider, organizations can cut down their time spent on setting up basics, thereby highly accelerate their development cycle in the initial stages [10] [12] [14].

(i) **Unlimited Storage**: Of course, there are still limits but storage capacity is almost unlimited in the cloud [10].

1.4.2 Market Demand for Cloud Computing

The market demand for the cloud computing can be summed up with the following points:

(a) Traditional IT infrastructure is proving very costly and time consuming to maintain for several enterprises. According to studies, an organization spends more than 70% of its IT budget in maintaining the previously setup IT infrastructure. Therefore, adding new capabilities to the existing system becomes a big challenge from monetary perspective. Many established enterprises are currently using cloud computing services due to this very reason. Big names like Coca-Cola, Mogulus TV channels, Animoto, Times Newspaper are in the list who was the early adopters of the cloud based services [6].

(b) There is an increased demand for storage and IT infrastructure from different enterprises. The demand is requires variable access and quick delivery Complete utilization of hardware resources is tough in case of distributed systems. We need eco-friendly systems that would consume less energy and allow more computation power [6].

1.4.3 Support of Government on Cloud Computing

(a) Government of several technologically advanced companies like U.S., Japan, Australia and U.K. are specifically encouraging the use of cloud computing. The use and demand of cloud computing is growing manifolds in the asia pacific region as well. Such a positive change is due to the fact that cloud computing all

the government and business firms to reap high benefits through the cost effective and powerful use of the information technology [16].

(b) Government of India is also very keen in adopting the cloud computing technologies. There are several e-governance projects and activities launched by the government throughout the country. In Indian manufacturing sector, the use of cloud computing has been prevalent for over two decades. As highlighted in several industry researches, since 2010, even the CIOs in the Indian manufacturing sector have actively adopted the cloud models [17].

(c) Technology giant Microsoft is going to invest over Rs. 1400 crores in the country to setup cloud data centers. Microsoft has also rolled out the "Cloud Accelerator Program" for the Indian businesses and government projects. Home-grown brand TCS is also setting up cloud data centers in the country. In fact, the acceptance of cloud based solutions and technologies like Big Data, IoT (Internet of Things) are the key indicators of IT growth in the country. It is being boosted further by the government initiatives for the "Digital India" [6].

1.5 CLOUD COMPUTING BUILDING BLOCKS

The building blocks of cloud computing paradigm constitute of three service models and four deployment models. These can be illustrated as below:

1.5.1 Cloud Computing Service Delivery Models

These are divided into three main categories which are:

A. Software as a Service, also called as SaaS.

B. Platform as a Service, also called as PaaS.

C. Infrastructure as a Service, also called as IaaS.

A. Software as a Service (SaaS)

In this, provider provides service to user for accessing the software to develop application where software is provided on monthly rent basis. More the user utilizes it the more he will be owed [83]. The Figure 1.3 representing Software as a Service (SAAS) model is shown below:

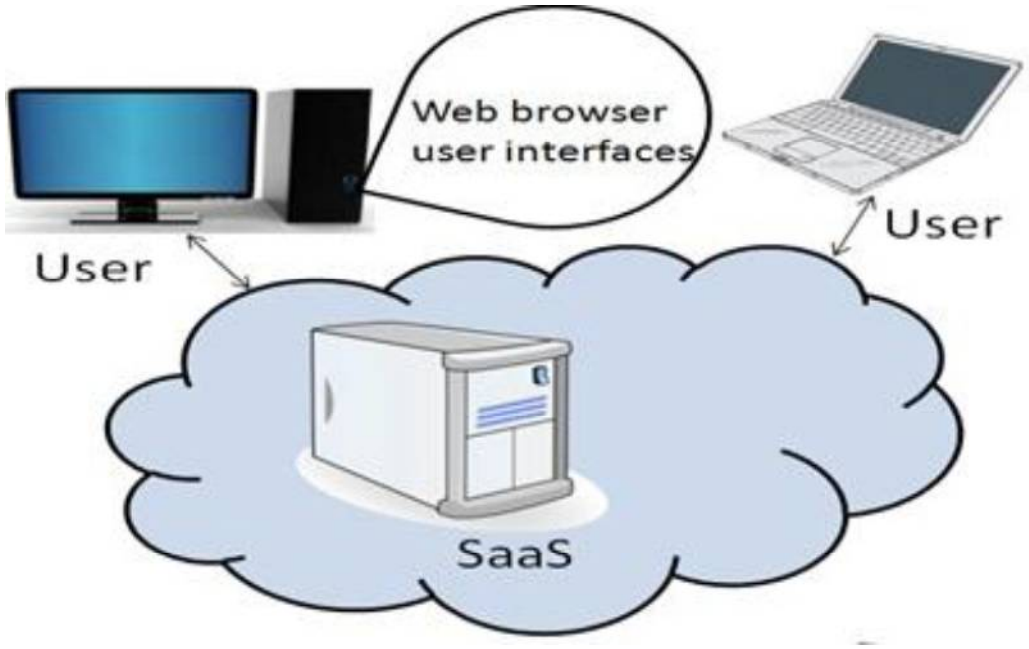

Figure 1.3 Software as a Service (SAAS) Model [8]

The Pros of Software as a Service (SaaS) can be described as:

(a) **Low initial setup cost:** Software as service applications are offered on a subscription based model. Hence, there is no need to purchase a license key and thereby, this helps reducing the initial setup costs for the software, hardware and even the manpower that's needed for the setup [18].

(b) **No need for upgrades by the custome**r: Traditionally, a business desired IT upgrades on a timely basis to keep up with the latest upcoming technologies. This was an unavoidable expenditure until the onset of services we are discussing [19]. In the SaaS model, the SaaS provide manages all the updates. Customer does not need to get into hassles of downloading and installing any patches [18].

B. Platform as a Service (PaaS)

PaaS was developed above the SaaS level and it provides software platform layer resources, with operating system support and software development frameworks on which user can build their own applications and host them. It provides all the resources needed to build an application [84]. The Figure 1.4 representing Platform as a Service (PAAS) model is shown below:

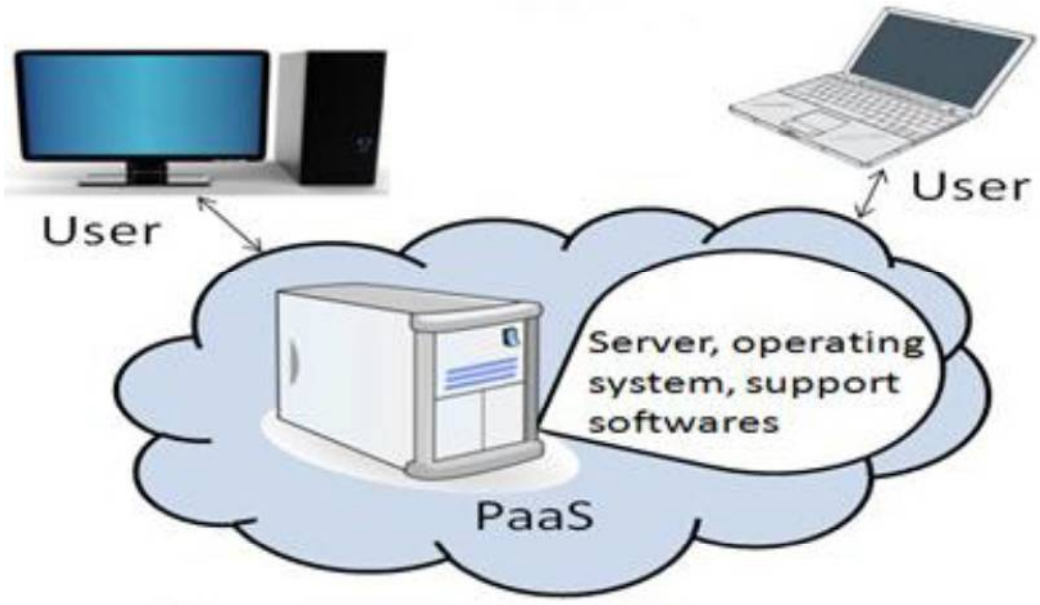

Figure 1.4 Platform as a Service (PAAS) Model [8]

(a) **Faster Development Rate:** As PaaS providers offer the end user pre-built functionalities through several APIs and user interfaces, the end user doesn't have to begin working from the scratch. As a result, development time is highly reduced thus benefitting the end user immensely [20].

(b) **Pre-built infrastructure:** As the infrastructure is handled by the service provider and is available via internet, the developers do not need to worry about the hosting, storage or any issue related to the basic infrastructure. They can directly focus on the main work i.e. development [20].

C. Infrastructure as a Service (IaaS)

IaaS [85] offers an infrastructure to the client. It allows clients with the access to server hardware, storage, bandwidth and other basic needs for computing resources. For example, Amazon EC2 allows individuals and businesses to rent machines preconfigured with selected operating systems to run their own applications. The Figure 1.5 representing Platform as a Service (PAAS) model is shown below:

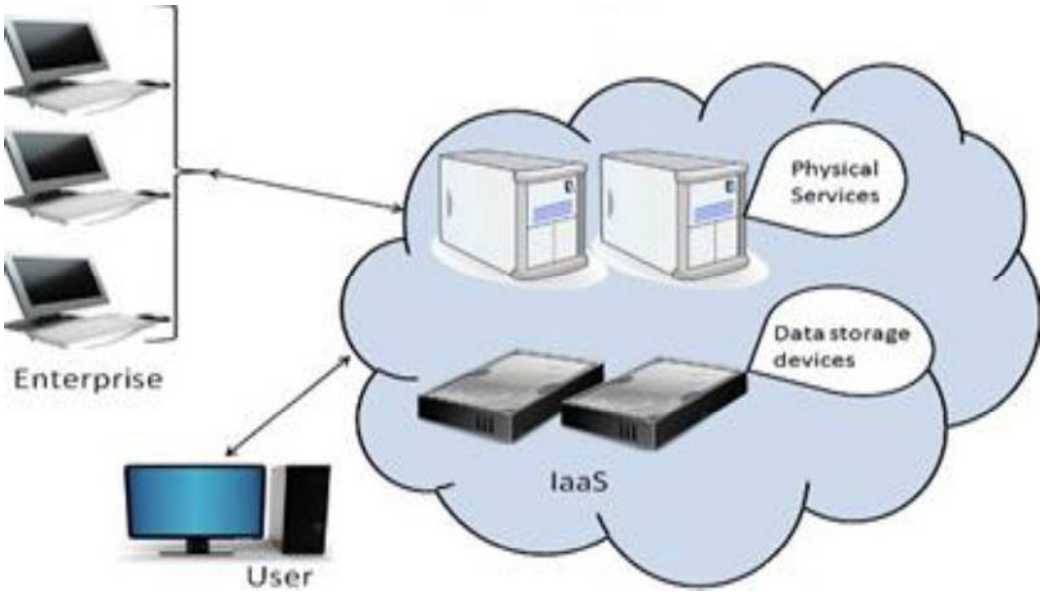

Figure 1.5 Infrastructure as a Service (IAAS) Model [8]

In IaaS client holds the control of storage, operating systems, deployed applications and perhaps constrained control of select systems administration parts. IaaS model is a low level of reflection that allows cloud clients to the right of the entrance the underlying foundation through providing proper access to virtual machines [8]. Automated administration of tasks, dynamic scaling etc is the few characteristics of IaaS.

In comparison with the various conventional service hosting environments like dedicated server farms it has been found that the architecture of cloud computing is more customizable. The layers are loosely coupled with the layers present above and below. The Figure 1.5 shows all the above cloud providers (SaaS, PaaS and IaaS) in different layers with the resources being managed. Each layer has its own importance in cloud environment.

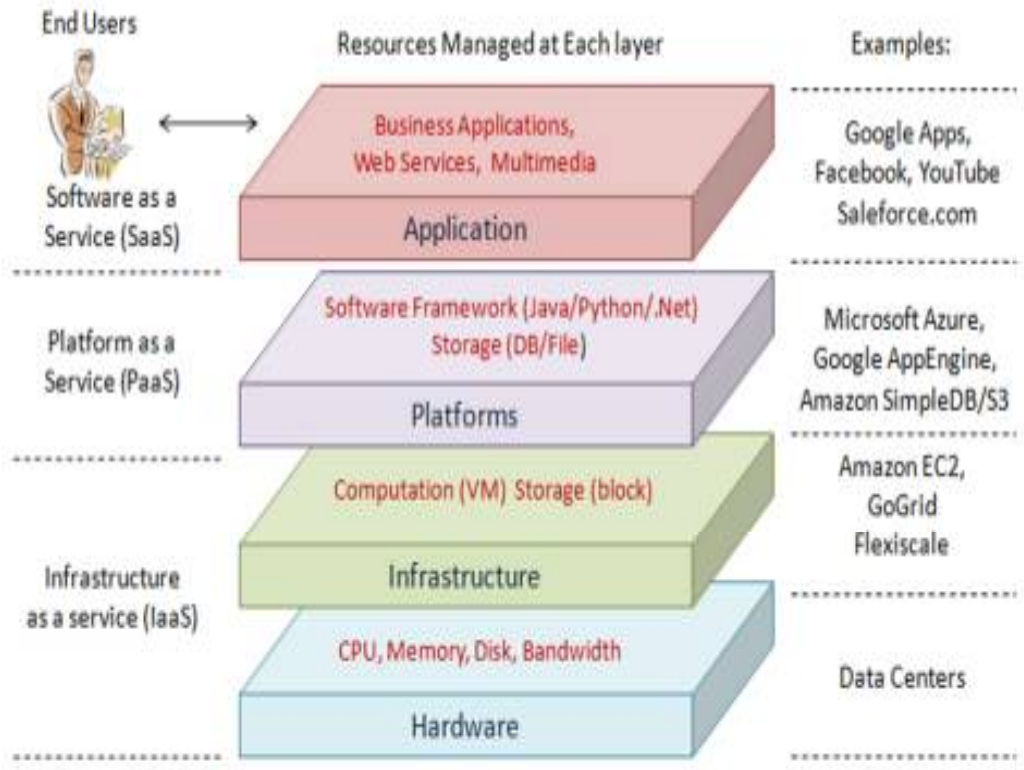

Figure 1.6 Layered Architecture of Cloud Services [22]

1.5.2 CLOUD COMPUTING DEVELOPMENT MODELS

Cloud services can be deployed in many ways, depending on the organizational structure and the provisioning location. The models deployed are [22]:

(a) **Public Cloud:** As obvious from the name, a public cloud is available to the end users - public or large organizations for use while the infrastructure is owned by a third party organization. It's hosted on the internet and can be used by any person who has access to internet and is paying for the services. The main advantages of public cloud include availability of resources anytime anywhere, continuous uptime, on-demand access and scalability, 24/7 support, easy setup and inexpensiveness. However, two major drawbacks compel serious users to avoid public clouds - data privacy and security [22] [119]. The Figure 1.7 represents public cloud architecture [120]

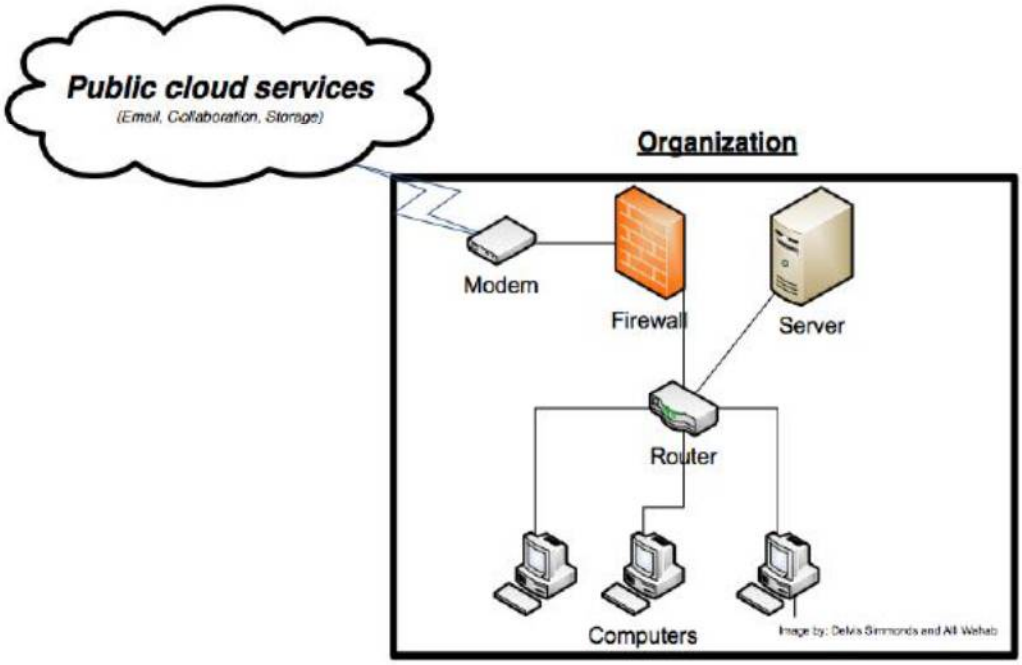

Figure 1.7 Public Cloud Architecture [120]

Public cloud computing offers various advantages like a greater scalability. Cost effective is one feature which makes it pocket friendly and saves money. Services like SaaS, (Paas), (Iaas) are easily available on Public Cloud platform which makes them easily accessible from anywhere through any Internet enabled devices.

Location plays no hurdle as it is location independent and the services are available wherever the client is located. Above all, it ensures reliability which means no single point of failure will interrupt your service reliable [128].

Apart from the various advantages there are some disadvantages of Public Cloud Computing which needs to be mentioned. First and foremost drawback is no control over privacy or security.

Secondly, a setback is that it cannot be used for use of sensitive applications. A major issue is that it lacks complete flexibility and the reason is as the platform depends on the platform provider. The data management protocols needs to be revised as at present protocols are not stringent [128].

(b) **Private Cloud:** As compared to public cloud, private cloud is more secure. In fact, that is the only big advantage a private cloud environment has over public clouds. However, with this add-on, it comes with a higher cost which makes it an unpopular choice amongst small-scale users working on tight budgets [22] [119]. The Figure 1.8 represents public cloud architecture:

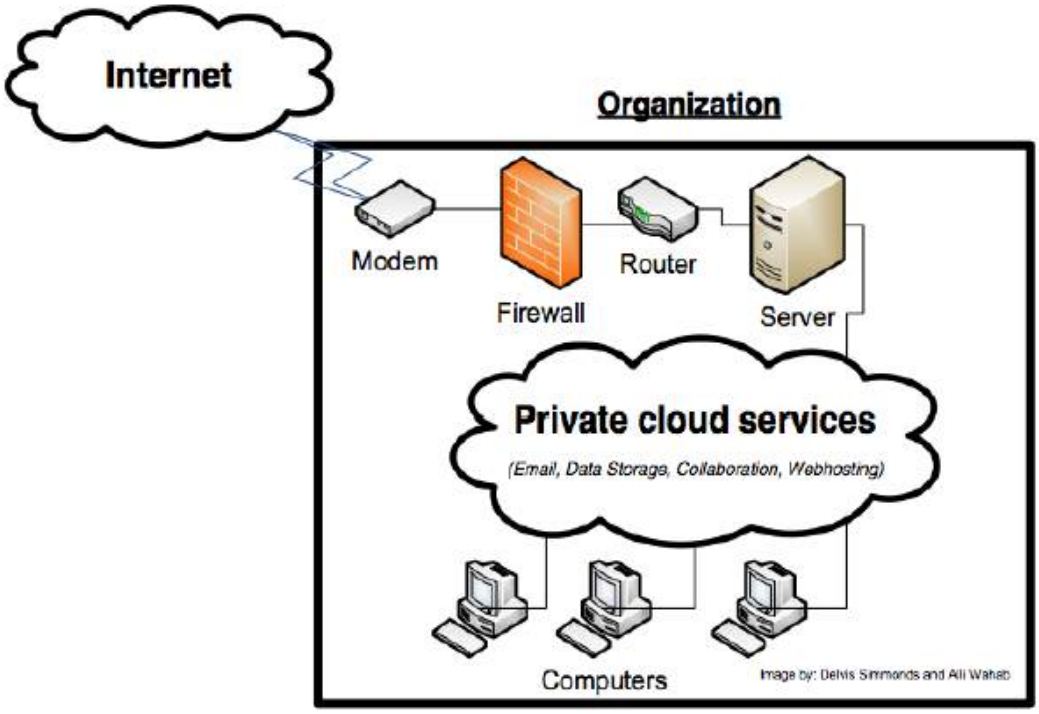

Figure 1.8 Private Cloud Architecture [120]

Private Cloud Computing also comes with pros and cons. The various issues fixed by private cloud computing are first greater security and privacy secondly it offers more control over system configuration as per the company's need and lastly greater reliability when it comes to performance. This feature enhances the quality of service offered by the clients. But all these features make it bit expensive when compared to public cloud and requirement of IT Expertise can be counted as a disadvantage [128].

(a) **Community Cloud:** A community cloud finds a middle ground between private and public clouds. Imagine it as a private cloud common to two or more organization its same set of security and privacy requirements. The resources are share within these organizations, thus bringing down the cost. The bandwidth and storage size remains fixed as they are shared amongst the community members. The Figure 1.9 represents community cloud architecture

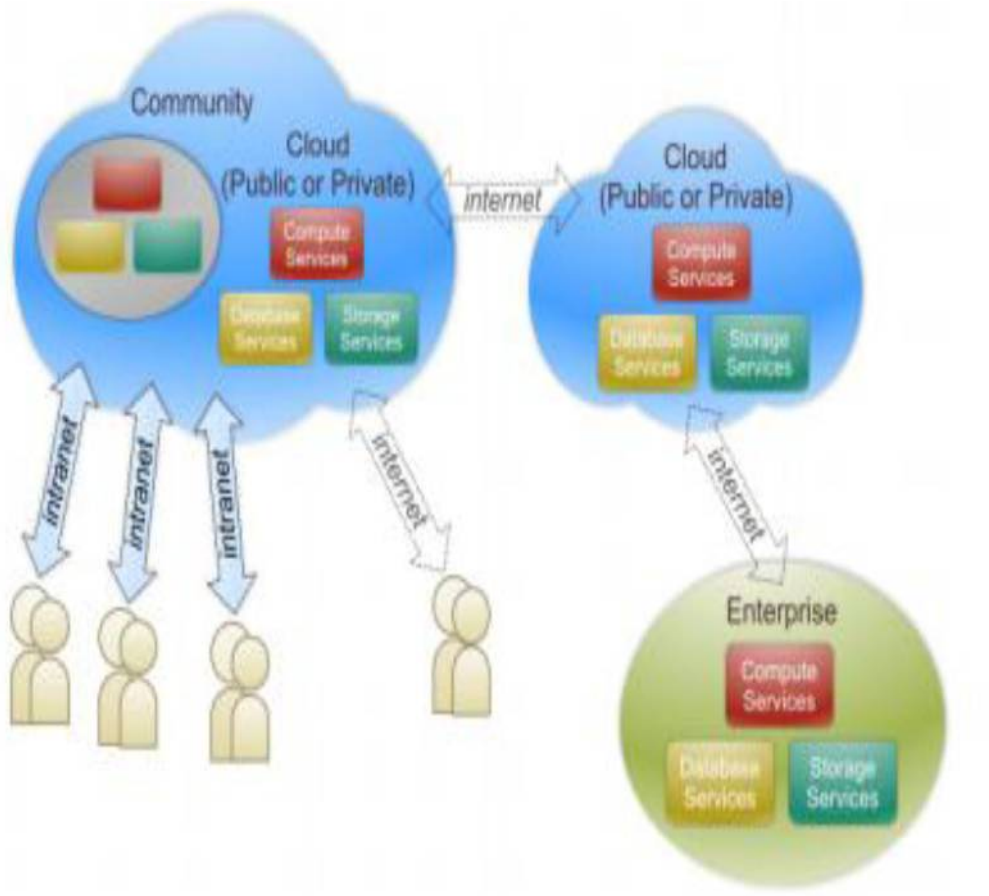

Figure 1.9 Community Cloud Architecture [122]

The third-party that offers a community cloud service is required to be impartial to all such organizations working together on sharing basis. That ensures fair allocation of resources. However, the cost of a community cloud still stays higher than that of a public cloud [22].

(b) **Hybrid Cloud:** As evident from the name, a hybrid cloud is a mix of private and public cloud. In the said composition, at least once private and public cloud must exist. Such a formation leads to reduction in capital expenses, improvement in resource allocation and overall agility. The drawbacks of such a system include a wider surface area for attackers to work on. The privacy and integrity are major causes of concerns in such an environment [22] [119]. The Figure 1.10 represents community cloud architecture

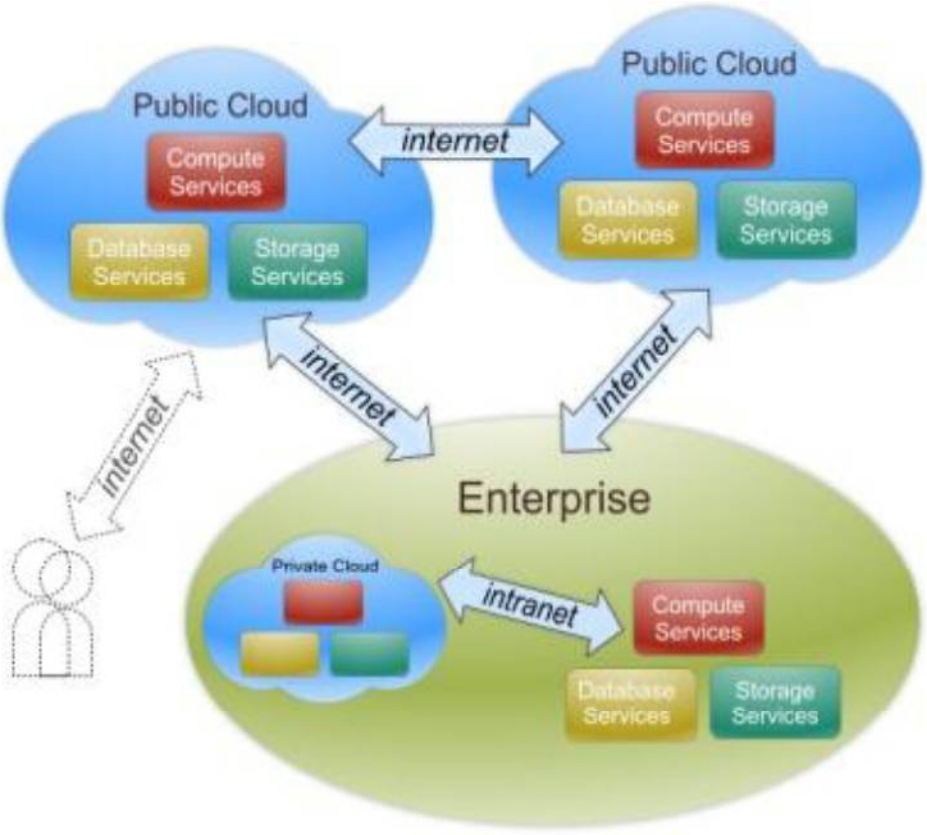

Figure 1.10 Hyrod Cloud Architecture [122]

Hybrid Cloud Computing features the same properties as private cloud computing i.e scalable, cost efficient, better security along with greater flexibility but Infrastructure dependency and possibility of security breach through public cloud are the shortcoming of hybrid cloud computing.

The below mentioned Figure 1.11 shows the major difference between the different types (Public, Private and the Hybrid) computing

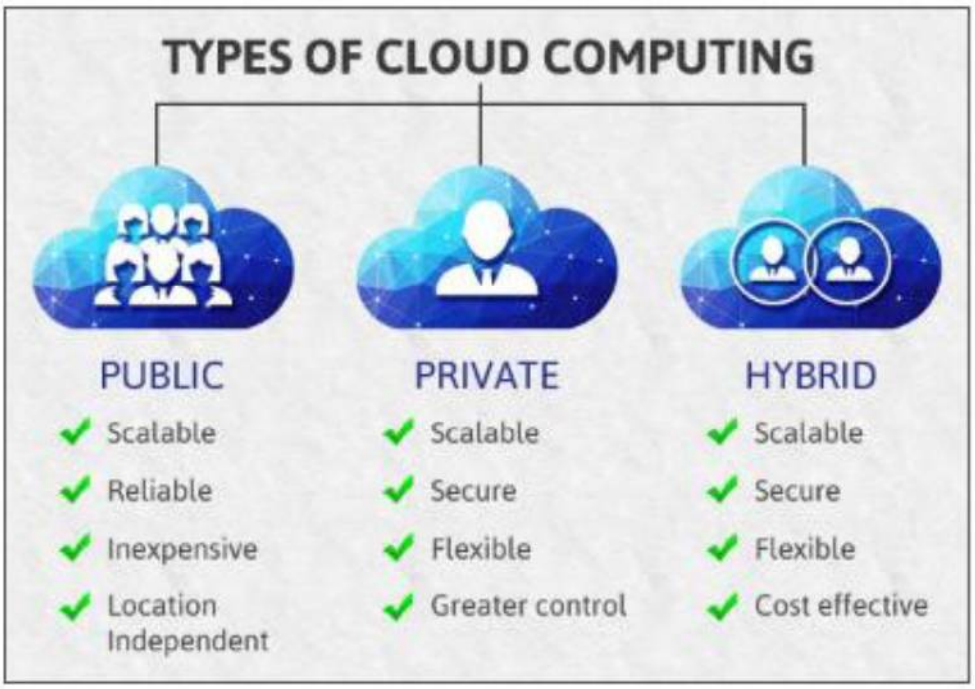

Figure 1.11 Differences among Types of Cloud Computing [128]

1.5.3 Cloud Computing Entities

There are two main cloud entities in the business market:

(a) **Cloud Service Providers**

(b) **Consumers**

There are two more service level cloud entities named as Service brokers and Resellers. All of these are elaborated below in detail:

A. **Cloud Service Providers**: Cloud Service Providers manage the cloud infrastructure and offer various services like SaaS, IaaS, PaaS etc to the other entities like customers, service brokers and resellers. Service brokers and resellers often end up playing the same role as cloud service providers in the following contexts [21]:

(a) Cloud Service Brokers: Cloud Service Brokers act as middlemen between the customer and cloud service providers. They help in negotiation between the two entities. They do not own or manage the infrastructure like cloud service providers do but they do add extra services on the top of what cloud service providers offer to the customers to make a good cloud environment for the customer/user [21].

(b) Cloud Resellers: Cloud Resellers are of utmost importance when a cloud service provider is looking for international business expansion. Any entity like a local IT consultation firm or a reseller of existing product can act as a "reseller" for the service in a specific region, as decided by the cloud service provider [21].

B. **Cloud Consumers**: Cloud Consumers are typically the end users of the cloud services. Sometimes, service brokers and resellers are also put in the same category when they are acting as a customer of some other broker or reseller [21].

1.6 APPLICATIONS OF CLOUD COMPUTING

Cloud computing due to the lost cost computational power and increase in flexibility, mobility, scalability and enhanced storage has become one of the most popular technologies at present time. The data produced by the users over the internet are stored in a remote location with respect to the user. The mentioned factors have made the cloud computing an active component in the application areas discussed below [122]:

(a) **E-Learning:** E-Learning is very popular in the education field. It offers a very effective and attractive solution for online learning without requiring any network and storage infrastructure at the user end. It facilitates students, research scholars and faculties to connect with the cloud over the internet and access the services and information online [122].

(b) **Enterprise Resource Planning:** Enterprise resource planning (ERP) is a business process management software that helps an organization to use integrated applications so as to manage business, services, applications, payrolls and various other branches of the company. It is highly cost ineffective for the organizations to have in-house ERP infrastructure. Cloud service providers install an ERP system in the cloud that can be used by multiple organizations without requiring any investment on the infrastructure [122].

(c) **E-Governance:** The service providers can easily improve government functions by working on the way it offers its services to the citizens. It also minimizes the overhead of installing, managing and upgrading the applications and services [122].

(d) **Backup:** Conventional backing up of data was done using tapes and drives. Storage at remote physical locations however, exposed the media to various kinds of accidental acts like robbery, natural calamity or accident etc which could lead to loss of data. The solution to this problem along with assured level of safety and isolation besides preventing accidental corruption of data has been provided by cloud based back up. Availability of cloud resources for storage of huge data with ease of transfer through wire and that too with Cloud based backup is certainly emerged as the most desired way of data back up and offered most feasible solution for both personal and business and enterprises [130].

(e) **Cloud Database:** Backend of web applications rely on data bases. Traditionally web developers used to manage and tune their databases besides huge money investments in databases. The Management of data bases involved a lot of complexities and required excellent skills. A feasible solution to this was cloud databases through which web developers were offered a chance to access, manage and tuned the databases for almost every requirement they indicate. One popular example in this category of services is Rack space [131].

1.7 CLOUD COMPUTING CHALLENGES AND BARRIERS RELATED TO ADOPTION IN AN ENTERPRISE

Following are the some of the research challenges posed by use of cloud computing which are centre of attraction for lots of researchers and presently tremendous amount of research is going on around the globe. These research challenges can be elaborated as follow [23]:

(a) **Handling the uncertainties:** It is one of the major difficulties in the provisioning of the cloud assets - uncertainty. Resource uncertainty can happen due to a number of issues like content type, heterogeneity client location and so on. Multimedia delivery apps can often suffer due to insufficient resources or inefficiency of shared resources. In multimedia apps, the bandwidth requirement depends on three main factors- application's bandwidth demand, bandwidth demand due to user workload and user location & device type. As these factors change quickly, the bandwidth shortage can often become an issue. One has to predict and provide enough bandwidth to resolve this issue [23]. Following factors play a crucial role in this:

Application's Bandwidth demand: Prediction of the size of the approaching video frames by using the transient history of the previous frames.

Bandwidth demand due to user workload: Prediction of user's workload i.e. number of people requisition a video by relying on the history of the previous user requests.

Resource Demand: Dynamic prediction of the relationship between a multimedia application, Quality of service targets, current hardware allotment, and the changes in a client's workload patterns [23].

By predicting and analysing above mentioned factors, a cloud resource provider can allocate proper amount of bandwidth without causing any QoS degradation.

(b) **Handling the dynamic variations in the workload:** An often unexplored benefit of the hybrid cloud infrastructure is its ability to handle peaks in the workload. It is possible to provision stipulate a local data-centre with appropriate serve ability to handle peaked workloads; by invoking the cloud assets to manage the peak demand. Such a methodology can help to arrange formats that can handle dynamic workloads [23].

(c) **Resource Allocation:** Various policies must be implemented for resource allocation in cloud computing based on service level agreement (SLA), centralized decision and distributed multiple criteria decision. The consumers' requirement is documented in the form if Service Level Agreement (SLAs) has been reached. Once an SLA has been established between client and service provider, the provider become liable to allocate the resources to meet the agreed level of service levels. This requires dynamic scale up or scale down of the resources to requests. Cloud computing vendors should survey that how different services contribute in resource allocation [87].

(d) **Energy Efficiency:** It involves efficient use of energy in the infrastructure, avoiding utilization of excess resources than actually required by the application and minimizing the carbon footprint of the Cloud application. Energy-aware job amidst scheduling and server consolidation are two other ways to reduce power consumption by turning off unused machines [88].

(e) **Efficiency of the virtual machines consolidation to manage heterogeneous workloads:** Cloud service providers allow the users to have the capacity to procure virtual machines and assign any kind of uses to them. It is possible that distinctive sort of uses will be allotted to the same physical machine node. It is

unsure how these applications will impact one another; thereby it's not evident whether there will be variable or static load on the resources. Thus, the issue is to find out what sort of users can be allocated to a single host and ensure proper utilization of resources without overloading them [23].

(f) **Quality of Service (QoS):** Cloud Service Providers (CSPs) need to ensure that adequate amount of resources are made available to fulfil the QoS requirements such as deadline, response time and budget constraints. These QoS requirements form the basis for Service Level Agreements (SLAs) and any violation will results in penalty. Therefore, CSPs need to make sure that these violations are circumvent or minimized by dynamically providing the specific amount of resources in a timely manner [132].

(f) **Security:** Confidentiality, privacy and integrity are most common challenges that may interfere with the potential power of the cloud computing paradigm. Confidentiality is the ability of an authorized group to access the protected data. As the number of users increase, the access points also increase. This sharp rise in access points makes the data more exposed to the external parties, thus making it more likely to get compromised. Privacy is the ability of a user or group of users to control the sharing of data with others. Individuals have to follow the rules set by the government regarding the privacy and confidentiality. But as the data is stored at multiple locations in the cloud, it is very vulnerable to the security breaches. Integrity is the means to protect the data, application and services from unauthorized access and modification in the cloud. It is divided in two categories first is data integrity, which is the protection of data from unauthorized modification. Cloud service provider must ensure no access or manipulation of the personal data of a user. The latter is software integrity as it refers to restricting the access and modification of the software applications in the cloud. Availability is the ability of the authorized user to access the cloud and use it for sharing information, resources and applications even if there is a malfunction in the system or a security issue. It includes the availability of data, physical components and applications on user request [123].

(g) **Scientific workflow scheduling:** A lot of scientific problems encountered in areas such as astronomy, physics and bioinformatics can be solved using cloud computing. But as the cloud computing faces performance variation challenges , which can affect the overall execution time of the workflow, and failure

challenges, which can affect the overall workflow execution as well as increase the execution time, it is a challenge to be achieved. The workflow scheduling on the distributed system is a widely studied subject and is considered NP-hard by reduction from the multiprocessor scheduling algorithms [24]. There are two main stages in the planning of the execution of a workflow in the cloud computing environment. The first stage is the resource provisioning phase in which the computing resources are selected and provisioned. The second stage involves the generation of a schedule and mapping of each task onto a best-suited resource is done. The selection of resources like this and mapping of the tasks is performed so that the different user-defined QoS (Quality of Service) requirements can be met [25].

1.7.1 Barriers Related To Adoption of Cloud Computing Technology in an Enterprise

Cloud computing offers a good opportunity for SMEs (Small and Medium business enterprises) to grow without worrying about the basic setup cost and enhancements in terms of innovation, computing functionalities, and the price [15].As with the increasing trend, small and medium sized enterprises (SME) are getting more inclined toward cloud computing because of the various strengths offer by this environment but there are various issues related to this technology that hinders several companies in adopting this technology. The below Figure 1.12 shows various concerns relevant to the enterprises:

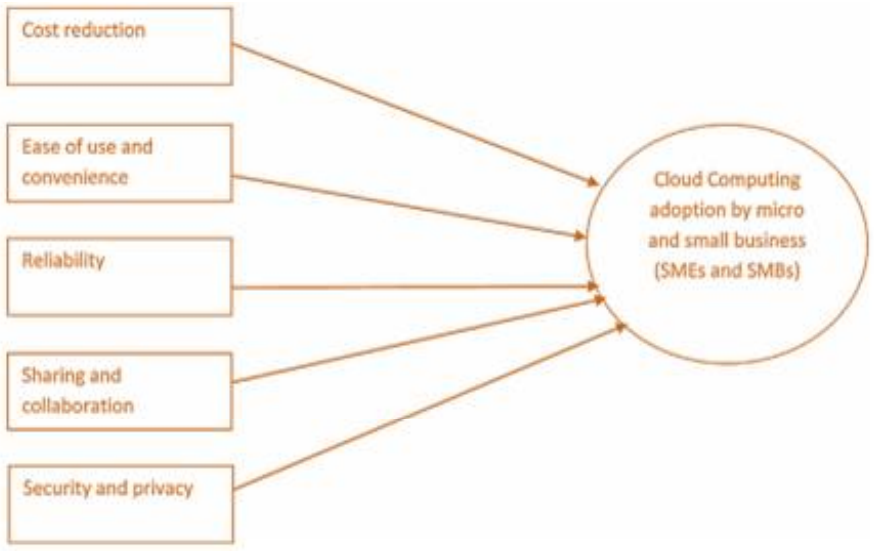

Figure 1.12 Cloud Computing Concerns for Enterprises [80]

(a) **Privacy & Security:** As cloud computing is still a fairly new computing model, there are uncertainties regarding the security at various cloud levels. This uncertainty has many executives reluctant of adopting a cloud computing model for their enterprise citing lack of clarity on security as a major reason. Also, cloud computing as a technology hasn't been able to address the privacy issues, they are questioned quite often. Nowadays, organizations face a lot of varied requirements regarding the protection of an individual's privacy and it's not yet clear whether a model like cloud computing can offer enough protection of information, or if an organization can actually use such a model without violating the security and privacy regulations[26] [27].

(b) **Connectivity Issues and Open Access Issues:** It is evident that the cloud computing can realize its full potential only if high-speed internet access is possible. Such kind of connectivity can globally open the possibility of a big industry and a lot of new range of consumer products, just like electricity availability did. This sort of connectivity as well as available open access to computing resources and information access through cloud promotes a new era of industrialization and there's need for more sophisticated products for the consumers [26].

(c) **Economic Value:** By measuring the return on investment, one can predict the growth of cloud computing. It's quite obvious that economic value will be there only by sharing of resources to level out the peaks, paying for what's used, and cutting upfront cost spent on the deployment of the IT solutions. All the costs and long term as well as short term benefits have to be carefully balanced. There are also hidden costs like disaster recovery, data loss insurance and application modification costs. Threshold values will be determined for which consolidation or combining the cloud resources would be feasible - economically and technically. For example, it won't be cost efficient to utilize more than one autonomous Software as a Service i.e. SaaS applications. Each will need to have a contract for the disaster recovery services. Also, economies of scale may mean that these functions must be combined into a likewise service [26].

(d) **Reliability:** Cloud servers portray the same problems as the resident servers do. There are downtimes and slowdowns in the cloud servers too. The only difference is that users have a higher dependency on the cloud service providers in the cloud computing model. But there's also the fact that once you choose a cloud service provider, you get locked in and that could be a big business risk [28].

(e) **Solution:** For proper advancement of cloud computing, proactive measures have to be taken to ensure security. It is crucial to evaluate access control, identity management, physical layer management, personnel layers management and the reporting of security incidents before you select a cloud service provider. Personal information should not be stored or sent through cloud as much as possible. A cloud service provider also should work on maximizing user control and offer paper feedback. Organizations must run the applications and transfer data in a private cloud and then, transmit it to a public cloud. There are several legal issues that exist in this type of computing model, but cloud security alliance needs to design standards to deal with such issues as soon as possible [28].

1.8 CLOUD COMPUTING VENDORS

Some of the companies researching in the domain of cloud paradigm are big names in the IT or computer industry. Choosing the right public cloud provider is becoming crucial because end user is not aware of the location of physical resources and device. The list for large cloud players in market holds the names of Google Cloud Platform, IBM Cloud and Microsoft Azure etc. but choosing among them totally depends on the enterprise's needs. In the web based application Google is considered the leader among the others vendors, so there is no doubt that the company offers excellent cloud user support and effectively dealing with the challenges being faced. The applications of cloud computing are fundamentally endless. Microsoft, IBM and Google are investing a huge amount of money vary from millions of dollars into the research [127]. The below mentioned Figure 1.13 shows some names of vendors in cloud computing

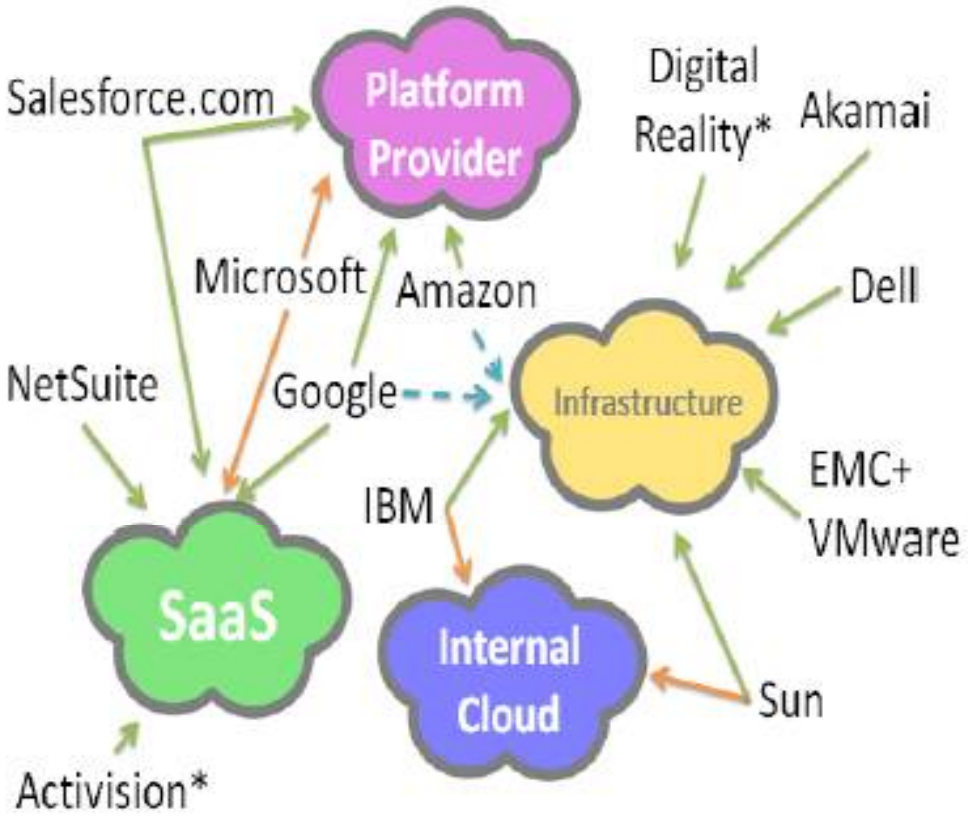

Figure 1.13 Vendors in Cloud Computing [127]

Here are some examples of cloud computing services provided by common cloud service providers:

A. Cloud Computing Services by Google: Google is known for integration of many apps and hence provides a number of cloud services to consumers. Such high-level integration makes Google a top CSP. It allows the consumers to get their tasks accomplished daily. It also helps consumers to save major amount of money as developing and maintaining software for such services and apps can be time consuming and very expensive as well. Here are some of the most popular cloud services offered by Google [29]:

(a) Gmail: Gmail is an email service that provides users with about 15GB of storage. Apart from storage, Gmail, effectively fights off spam and allows mobile access too. It also has a chat applet called Hangouts that can store conversations in the form of email. Google also offers business packages for services like Gmail under Gsuite, allowing many small and big enterprises to optimize their business communications via email.

(b) Google Docs: Google Docs is emerging as the biggest contender for desktop applications like MS Excel, MS Word and MS PowerPoint. Google has launched their counterpart under Google Docs so that users can make spreadsheets, presentations and craft word documents, all online and store them in the cloud services. Therefore, such documents are always available online and can be accessed from anywhere and at any anytime. It is highly convenient for team members that are working remotely. They can co-operate and work on the same document without any delay due to the distance. Such documents are very secure as well because the files are encrypted using a very advanced encryption technology. They can be accessed only by the authorized users.

(c) Google Analytics: Google Analytics is a service that helps to monitor a website's traffic. By doing a one-time setup, all visitors that come on a website can be tracked. It is highly detailed tracking allowing you to see traffic demographics and sources. There is also advanced e-commerce tracking for e-commerce businesses where goals can be setup and analyzed to keep track of sales and performance of the website.

(d) Google Ad words and Google Ad sense: Both of the above mentioned i.e. Google Ad words and Ad sense are advertising tools. Ad words can be used by businesses to run several ads like search campaigns, display campaigns and dynamic remarketing campaigns. Ad sense is a service used by bloggers and websites to display ads on their website. Google shares a part of their advertising revenue with such Ad sense publishers. Both are highly popular products by the company.

(e) Picasa: It had been one of the most popular products by the company but Google has now stopped this service.

This is not a complete list of cloud services offered by Google as the company keeps launching more products time to time. The above mentioned are however the most widely used and successful ones by the company.

B. Cloud Computing Services by Microsoft: The cloud platform offered by Microsoft is also called Windows Azure. It consists of a set of cloud services that are offered mostly to application developers and users. All the services are run in the

Microsoft Data centres that have setup all around the world. These services include [30]:

(a) Windows Azure: Windows azure is a windows environment that can store data and run applications in the cloud.

(b) Windows Azure App Fabric: This app provides infrastructure for the applications that run in the cloud or within an organization.

(c) Windows Azure Marketplace: Windows Azure Marketplace is an online market dealing in application software and data.

(d) SQL Azure: SQL Azure consists of relational database services that deploy a specific version of Microsoft SQL service for its functioning.

C. Amazon Web Services: Amazon Web Services offers solution for business of all sizes. Thanks to AWS, companies can provision cost effective, flexible, and scalable IT infrastructure and services. The services can also be scaled down whenever required. AWS allows companies to select a platform that would be suitable for their specific needs and companies, staying true to the cloud computing model, pay for only what they use. On top of that, AWS also ensures top level security by using advanced data privacy and physical security techniques to protect user data. The comprehensive AWS platform offers the following services [30] [31]:

(a) Amazon Simple Storage Services (Amazon S3): Amazon S3 offers secure, reliable and scalable storage for internet. It can be used to store & retrieve data of any size on the web.

(b) Amazon Route S3**:** Amazon Route S3 is a scalable and reliable DNS service.

(c) Amazon Simple DB**:** Amazon Simple DB offers core database functions.

(d) Amazon Virtual Private Cloud (VPC): Amazon Virtual Private Cloud (VPC) helps to connect the existing IT infrastructure of a company to the Amazon Web Services cloud through a Virtual Private Network (VPN).

(e) Amazon Elastic Compute Cloud (EC2): Amazon EC2 is a web service that offer configurable cloud computing resources.

(f) Amazon Cloud Front**:** Amazon Cloud Front is a web service that offers content delivery. It can transfer customer data with high speed and minimal delay through use of its global network.

(g) Amazon Relational database services (RDS): Amazon RDS is a web server that can manage a relational database in the cloud.

(h) Amazon Simple Queue Service (SQS): Amazon SQS is a reliable, scaled and hosted queue to store messages.

D. Cloud Computing Services by Salesforce.com: Salesforce.com has already made a significant mark in cloud computing development. It is probably best known for its sales management SaaS, in fact leading in this domain. The company's cloud computing architecture is dubbed Force.com. This highly desired and demand service platform running across the Internet. Sales force provides its own Force.com. Application Programming Interface (API) and developer's toolkit pricing is on a per log-in basis. Developers can have a multiple advantages like can use App Exchange applications uploaded by others, share their own applications in the directory or publish private applications accessible only by authorized companies or clients. Although many application in the App Exchange library are free, and others can be purchased from the original developers [127].

1.9 PERFORMANCE INDICATOR

(a) **Job Scheduling:** Job scheduling is a very important and demanding issue in the cloud computing system. It can be simply viewed as just a set of jobs that need to be assigned to various virtual systems. But it's so much more complex as it needs to consider several users defined and provider defined factors like deadlines, make span, energy and cost efficiency etc. Schedulers have to ideally map the user tasks to the virtual machines while satisfying the requirements and constraints put forward by both users and cloud providers. As the number of tasks submitted to the scheduler is also huge, the problem becomes even more complex. Most of the applications in the cloud environment can be divided into three types of jobs - dependent, independent and parallel. A task is actually the

interdependence of these jobs. Once an application is submitted to the cloud, it is broken up into several tasks. A scheduler decides upon what resources have to be allocated to these set of tasks and figures the execution order for these as well. It also has to manage any overhead that is related to the virtual machines.

Job scheduling is a very fundamental issue in the cloud computing environment. It aims to seek out responsible Virtual Machines for the tasks. Any job may contain a series of tasks and all such tasks are computational and independent of one another. Task scheduling is also a big challenge that impacts the efficiency of a scheduling algorithm, its design and implementation is highly effected due to this [35].

Various parameters are used to perform job scheduling, increasing the overall performance of the cloud. A task may involve entering of data, accessing software or storage or processing data. The data center works on classifying these tasks according the services requested and service-level agreement. Then, each task, classified by data centers, is assigned to any one of the available servers. The servers, then, perform the requested task and transmit the response to the user.

(b) **Energy Efficiency:** In today time the cloud infrastructures have recently becoming a centre of attention around the globe. In cloud environment, to boost the return on investment, cloud providers have to implement energy-efficient job scheduling strategies. In cloud computing the scheduling framework which is targeting a reduction of energy consumption must focus on the implementation of the optimal scheduling algorithm and perform the allocation of cloud user's jobs on the different virtual machines efficiently.

One of the primary causes of energy inefficiency in cloud environment is the idle wasted when scheduler not properly schedules the users jobs to virtual systems which result in low utilization of virtual resources. Hence, energy-efficient job scheduling solutions are needed that can address the high increase in the energy consumption from the perspective for cloud provider. Heuristics algorithms assign each cloud user job to virtual resources on which the energy consumption for executing the jobs is minimized without any performance degradation.

1.10 RESEARCH OBJECTIVES

Cloud computing, as a technology, is undergoing consistent evolution and hence, confronts several challenges day by day. However, scheduling is one of the challenges faced by cloud computing that's often neglected, and it totally needs more work for the overall evolution of the technology in real-world. Scheduling is defined as a set of constructs that are made to control the order of work that has to be performed by a pool of virtual systems in cloud. In cloud environment, the job scheduling framework has a big impact on the overall performance. The scheduler is responsible to select the user's jobs from the available job pool, and then map it to the resources available, that's why scheduling is considered as an NP-Hard i.e. non-deterministic polynomial time problem. For e.g., if the scheduler has to map 40 jobs to ten different resources, it will have about 40*10 possible solutions which signifies that any job can be mapped into any resources. Another complex issue faced by a scheduler is that it almost never has accurate information about the available resources that can deal with a given set of jobs. In particular, in this research the following research problems are investigated:

(a) Study and performance analysis of various existing scheduling algorithms in terms of energy efficiency.

(b) Design an energy efficient heuristic approach for scheduling data-intensive applications in cloud computing.

(c) Develop a priority based energy efficient scheduling technique for optimal job processing for energy saving in cloud environment.

(d) To purpose and implement a scheduling policy considering the tradeoffs between performance improvement and energy saving.

The aim of the thesis is to make the whole cloud environment more reliable and effective by proposing enhanced job scheduling technique. This thesis will focus on to make the more efficient job scheduling framework in cloud computing by developing a well-optimized energy efficient job scheduling algorithm to counter the above concerns.

1.11 CONTRIBUTIONS

All the information and services provided via cloud computing is shared on the internet, that is why it is crucial to understand various issues associated with it. Indicating the challenges in cloud environment and assessment of existing works uncover gaps in performance measures. Considering that Scheduling of cloud user jobs in an effective job scheduling framework which can provide energy saving that have direct impact on the performance of clouds; we addressed several issues related to performance in cloud environment.

1.11.1 Motivations

(a) **Job Scheduling in Cloud Computing**: The prime objective of job scheduling is to decrease the under utilizations of the prominent cloud infrastructure thereby providing the effective efficiency. The cloud paradigm hold the pool of resources on which several user jobs get executed and storage of these resources is huge and to achieve high performance in such scenario, scheduler implement scheduling algorithm that carry out jobs allocation in such a way as it can provide efficient job scheduling framework.

(b) **Energy Saving in Cloud Environment**: Scheduling is widely used as an effective energy conservation method in cloud computing paradigm as cloud environment provides the access to single server for multiple cloud users. So users can retrieve and update required data, as executing submitted jobs on maximum resource of every server directly impacts the energy usage of the cloud infrastructure in cloud data centres.

(c) **Heuristic Based Job Scheduling**: Most of the traditional job scheduling algorithms leave several issues unaddressed. Only few well heuristic algorithm which target the energy saving are proposed and addressed in literature. Most of the traditional algorithms gets failed to find the optimal execution schedule of submitted cloud user jobs. The primary challenge of scheduling in a cloud environment is the allocation of available resources effectively thereby improving efficiency of the whole cloud computing environment. Most Heuristic algorithm not considers resource availability in its scheduling decision.

1.11.2 Overview of Contribution

The research contribution of this thesis is elaborated as follows:

(a) Study of several job scheduling algorithms in cloud environment and describes their characteristics and features, advantages and disadvantages of each of them. First the relevant literature specific to cloud environment was thoroughly studied. Then the job scheduling in clouds is elaborated with the deep study of tradition and heuristic based job scheduling algorithm. This helped in identifying two main heuristic algorithms i.e. Efficient Multi Queue Scheduling (EMQS) Algorithm and Ant Colony Optimization (ACO) Algorithm.

(b) Secondly, Implementation of the efficient multi queue job scheduling algorithm (EMQS) algorithm and ant colony optimization algorithm (ACO) These algorithms were implemented by setting up cloud environment . The results of both algorithms are analyzed, compared and discussed on the basis on performance metrics of energy consumption and execution time.

(c) So keeping in mind the issue faced with Efficient multi queue job scheduling algorithm (EMQS) Algorithm, in this research work we worked in developing a improve job scheduling heuristic algorithm for cloud networks and named it as smarter multi queue job scheduling algorithm (SMQS) which focused on providing energy saving solution by making optimize utilization of all resources for execution of cloud user jobs as it considers the status of all virtual machines at the time of allocating user jobs.

(d) Lastly, as proposed SMQS (Smarter Multi Queue Scheduling) algorithm was formulated using two main criteria as first it divides user's jobs in shorter number of queues and the scheduler do the jobs allocation in such a way that it can provide energy efficient job scheduling framework. SMQS algorithm was then implemented in cloud environment, evaluated and compared with EMQS algorithm for scheduling cloud user jobs. The results showed that proposed smarter multi queue scheduling algorithm show rise in the performance of cloud network by sharp reduction in energy consumption and decreasing execution time to a certain degree.

1.12 ORGANIZATION OF THESIS

Research work of this dissertation has been organized into 7 chapters. Chapter 1 introduces cloud computing, its challenges and applications. It also covers the economic and technical advantages related to cloud computing. Chapter 2 presents a detailed literature review on cloud computing, job scheduling and energy efficiency. Chapter 3 presents a various taxonomies related to job scheduling. Chapter 4 discussed about energy efficient heuristic approach for scheduling data intensive applications in cloud computing. Chapter 5 explores the formation of priority based heuristic based energy efficient scheduling technique for optimal job processing for cloud environment carried out during this research work. Chapter 6 represents pseudo code of implementation and comparison results of proposed approach with the already exiting approach. Finally Chapter 7 presents the conclusion and future aspects of this research work.

1.13 SUMMARY

This Chapter gives the detail introduction about cloud computing highlighting the facts what makes cloud computing different from the other technologies. The evolution and technical aspects about cloud environment are discussed and prominent building blocks for clouds are also elaborated. Further research challenges which are pertaining in the area of clouds are also represented. At the end this Chapter also covers research objectives and methodology for carrying out the research work.