International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Enhancing Scalability in Cloud Computing: Strategies and Best Practices

Sai Tarun Kaniganti

Abstract: ***Computer security issues exacerbate with growth of the Internet as more people and computers join the web, opening new ways to compromise an ever - increasing amount of information and potential for damages. However, an even bigger challenge that has been created with the implementation of Cloud Computing. This chapter gives a description of information security issues and solutions. Some information security challenges that are specific to Cloud Computing are described. Security solutions must make a trade - off between the amount of security and the level of performance cost. The key thesis of this chapter is that scalability applied to Cloud Computing must span multiple levels and across functions. A few key challenges related to Cloud Computing and virtualization are presented. Our goal is to spur further discussion on the evolving usage models for Cloud Computing***

Keywords: Cloud scalability, Scalability issues, Scalability Factor, Scalability Services, Horizontal Scalability, Vertical Scalability, Auto Scaling, Auto Scaling techniques, Auto Scaling Algorithms

1. Introduction

Scalability is a frequently - claimed attribute of multiprocessor systems. While the basic notion is intuitive, scalability has no generally - accepted definition. For this reason, current use of the term adds more to marketing potential than technical insight. In this paper, we will examine formal definitions of scalability, but I fail to lind a useful, rigorous definition of it. I then question whether scalability is useful and conclude by challenging the technical community to either (1) rigorously define scalability or (2) stop using it to describe systems. Cloud computing has revolutionized the way businesses operate, offering scalable and flexible solutions for their computing needs. Scalability, in particular, is a crucial aspect of cloud computing, enabling organizations to dynamically adjust their resources based on fluctuating demands. This research paper explores the concept of scalability in cloud computing, its importance, and its practical applications in various industries. Cloud Computing is effectively powerful computing paradigm to deliver services over the internet. It is a model to facilitate or enable on - demand network access, on - demand self - service convenient to a shared pool of computing resources in configurable manner which can be quickly provisioned. The model of Cloud Computing has been differentiated into IaaS (infrastructure - as - a - service), SaaS (software - as - a - service) and Paas (Platform - as - a - service). Cloud Storage is a service, in which the data is remotely managed, backed up, maintained and restore and it makes data available to users via internet or network. Several cloud storage providers provide free space up to certain gigabytes. For example, Drop Box offer free space up to 2GB, Google Drive, Amazon, Apple Cloud make available free space up to 5GB, Microsoft Sky - Drive give free space up to 10 GB

This paper is a first step to building that understanding. As a primer for policymakers on the cloud, this study outlines how to conceptualize the cloud and describes the evolution of the cloud market.

It then discusses cloud security in detail, using a timeline of past incidents together with in - depth case studies of the most significant incidents that are publicly known. Together, these serve as a foundation for developing a comprehensive framework for mapping the various risks and a severity schema to prioritize them. The paper then briefly outlines additional public policy issues to take into account while considering cloud security. Finally, it sums up and discusses the implications for public policy, while listing promising areas for future security - related research. Cloud computing philosophy enables users to use computing resources on rent basis. Unlike traditional computing, users need not to invest huge capital at the beginning for developing computing infrastructure and can pay to the cloud vendors as per their usage of resources. Resource consumption is measured and billed. When use is higher, users pay accordingly and during low usage they pay less. This is why cloud computing is called a variable cost computing model.

Origin of the term 'cloud computing'

The origin of the term 'cloud computing' dates back to the early 1990s. In those early days of network design, network engineers used to draw network diagrams representing different devices and connections among them. In such diagrams, they used to represent outer network arenas with cloud symbol since those details were not in their knowledge. This was known as 'network cloud' or 'cloud' in the networking industry during that period, but today we do not mean 'cloud computing' in the same sense. With the beginning of utility computing initiatives towards the end of the last century, major software firms focussed on deliver applications over the Internet. Email services gained pace during this period as the vendors started to offer the facility to their users. And the most remarkable initiative came from Salesforce. com when they delivered business application for enterprises over the Internet in 1999. But all of these efforts were seen as part of utility computing facility development. Cloud computing did not emerge till then.

The initiative from Salesforce. com in the year of 1999 to deliver business (enterprise) applications via a 'normal' website is considered as the first - of - its - kind effort. The success of Salesforce's effort encouraged other software firms to deliver business applications via Internet. That was the breaking point, when computing technology firms started to take initiatives in developing cloud computing based business applications. Salesforce. com launched in 1999 was the first successful commercial initiative to deliver enterprise

applications over the Internet. This was the first step towards cloud computing. Next major initiative came from Amazon with launching of Amazon Web Service (AWS) in 2002 which delivered computing services over the Internet. AWS provided a suite of services including storage. Amazon played the key role in the development of utility model - based computing services during that period, and soon many software firms started to modernize their data centers to support utility computing. A data center is an organized physical repository of computing systems and associated components, like storage and networking, which enterprises maintain to build their computing facility. The movement slowly turned towards what is known as cloud computing today. In 2006, Amazon launched its EC2 web service, where companies and individuals could rent (virtual) computers for running their own computer applications. Soon after EC2 started attracting the attention of experts, 'Google Docs' introduced by Google (in 2006) brought cloud computing service to the public attention. In 2007, Salesforce. com launched another service called force. com, where anyone could build applications and launch websites.2009 saw Microsoft's commercial entry into cloud computing arena with the launch of 'Windows Azure'. The goal of Azure was to provide customers the facility of running their Windows applications over the Internet. Apart from these commercial initiatives, many research organizations and open - source forums started their cloud computing initiatives during those years. For instance, NASA developed and launched an open - source cloud computing platform called as 'Nebula' in 2008 for its internal use

What is the cloud?

At its most basic level, the cloud is simply someone else's more powerful computer that does work for others. There is no one single cloud—so while it might be accurate to say that data crosses the internet, it is not correct to say that such data is stored in an ephemeral form, hovering somewhere in the sky. In fact, the cloud stores and transports data across a global infrastructure of data centers and networks. A more accurate description of the cloud is that cloud services are an abstraction of a parallel system of computers, data centers, cables, infrastructure, and networks that provides the power to run modern enterprises' and organizations' digital operations and to store their data. Building the necessary infrastructure for cloud services on a truly global scale has been one of the most significant architectural achievements of the past decade—and it mostly exists behind the scenes, out of common knowledge. With that said, as chapter 2 highlights, the cloud marketplace has evolved significantly over the years, as has the cloud itself.

To make sense of the transformative impact of cloud services, first consider how computing, for example, worked prior to widespread cloud adoption. In the past few decades, for every computational task that a company or individual needed to do, they had to have their own computers, servers, and even data centers. For instance, Capital One, a major company in the financial services sector, announced in 2015 that it would move all of its apps to the AWS cloud, meaning that it subsequently did not have to build and buy data center storage as it rolled out new apps. For smaller businesses, the costs of information technology (IT) procurement—that is, buying all the necessary computers and setting up the necessary networking for inhouse data storage and processing capabilities - were prohibitive to rapid growth. When companies like Amazon, Google, and Microsoft began to offer storage and computing power as services in the late 2000s, they changed this paradigm. These massive IT giants could manage networks of data centers, servers, and networking at global scales—meaning they could take advantage of economies of scale to offer computing as a service—at prices that would beat internal costs for most companies and still make them a profit, especially after significant price drops starting in 2014. Amazon, Google, and Microsoft particularly focus on providing the basic elements of IT infrastructure—server space and computing power—that are highly scalable, custom - configurable, and capable of being rapidly deployed and shifted. However, cloud computing encompasses a wide range of service types in which different firms predominate (see chapter 2), and the services provided include the basic infrastructure to build digital platforms on top of ready - made applications delivered over the internet. These various services can be grouped into the three principal types of cloud services. In practice, the major CSPs offer different services spanning all three of these categories.17

- Infrastructure as a service (IaaS): CSPs provide basic access to storage, networking, servers, or other computing resources.
- Platform as a service (PaaS): CSPs provide an environment—a platform—for customers to build and deliver applications.
- Software as a service (SaaS): CSPs build, run, and host applications delivered over the internet, which customers pay to access.

Defining Cloud Computing

Given the particular importance of cloud computing as a service, it is worth considering a 2011 definition of cloud computing by the U. S. Department of Commerce's National Institute of Standards and Technology (NIST) —the agency that sets technology standards. According to NIST, "cloud computing is a model for enabling ubiquitous, convenient, on - demand network access to a shared pool of configurable computing resources (. . . networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. " 19 Essentially, this definition touches on five key characteristics of cloud computing: 1) on - demand self - service, 2) rapid elasticity, 3) measured service, 4) broad network access, and 5) resource pooling. The first, on - demand self - service, means that customers can use capabilities only when needed and don't have to pay more. They are also siloed from each other even while making use of the same resources. customer selects additional computing capacity. Automatic digital systems handle the allocation, provisioning, and deployment of the needed infrastructure, platforms, and services.

Second, rapid elasticity means that the amount of resources dedicated to any one customer can at any time quickly increase or decrease depending on the needs of the customer. Because the resource capacity of a CSP is exponentially larger than the likely needs of any one customer, customers can scale their operations rapidly without taxing the CSP.

Third, measured service captures how CSPs manage and price their services. CSPs, like AWS and Microsoft Azure, charge customers for the resources they use on a unit - per - time basis— these units are an abstraction of the resources used. For instance, AWS's Elastic Compute Cloud measures its service in units that AWS defines n select computing capabilities automatically—without needing any human support from the CSPs they use. And on the CSP side, on - demand self - service means that any customer request can be handled automatically. No technician has to go configure a server when a in terms of standard central processing unit (CPU) integer processing power.

Fourth, broad network access means that customers access these services over the network, including potentially the public internet. This is a straightforward but highly important point from a security point of view. No longer are computing resources solely part of a firm's internal network—instead, in many cases, the core operating systems rely on connections that could be open to the entire global internet. In this respect, both the capacity and security of these network connections are critical. Even private cloud solutions, where the cloud servers are accessed over a private connection, would require significant bandwidth.

Fifth, resource pooling means that a CSP combines its resources such that each customer shares the same infrastructure with other customers in a dynamic fashion, to be apportioned and reapportioned as necessary. This feature is what makes cloud computing a more efficient model than separate computing resources for each firm. CSPs can take advantage of economies of scale to build infra - structure and platforms at mass scale—and share these resources among multiple customers at the same time to save costs and unneeded capacity.

Underlying Technologies While modern CSPs rely on highly complex systems for allocating, managing, and deploying re - sources among millions of customers, three key technologies are essential for understanding how cloud services work at scale: virtualization, hypervisors, and containerization. Virtualization allows for abstraction between physical hardware and individual computers.

Essentially, virtualization allows for multiple computers, referred to as virtual machines, to exist on the same physical server. Beyond computational tasks and storage, entire networks can be built through virtualization. Hypervisors are programs that manage virtual machines, servers, the connection between those virtual machines and servers, and the allocation of resources to the virtual machines. Thus, it be - comes possible to have a whole bank of physical servers, each running a hypervisor, and then create virtual machines across this bank of servers. CSPs refer to the virtual machines they create for their customers as instances, since they only last as long as needed and, once they are no longer needed, they are spun down to free up capacity. Containerization is a refinement of virtualization that works by running discrete containers within the same operating system, basically moving up the abstraction provided by virtualization by one level. Containerization caught on around 2014 with the introduction of a new tool called Docker, which made it much more convenient and efficient to implement containerization for business uses.

While a virtual machine includes an entirely virtual operating system, a container is an isolated environment within one single operating system. In terms of layers, while a hypervisor lies between the hardware and virtual machines, each with their own operating system inside them, a container sits on top of an operating system that is on top of the container engine, and then the hardware is below.

the infrastructure for the cloud is visualized according to a layer's model, similar to the Open Systems Interconnection model for the internet itself.22 The table presents a hierarchy of layers from the physical data centers at the bottom to the entirely virtual application layer at the top, allowing the various parts of the cloud to be simplified.

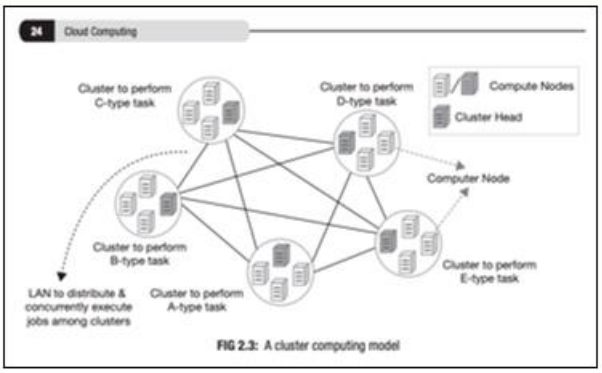

FIG 2.3: A cluster computing model

Scalability in Cloud Computing

Scalability refers to the ability of a system to handle increasing or decreasing workloads by provisioning or de - provisioning resources accordingly. In the context of cloud computing, scalability is achieved through the use of virtualization and resource pooling, allowing users to scale their computing resources up or down as needed.

There are two main types of scalability in cloud computing:

1) Vertical Scaling: This involves increasing or decreasing the resources (such as CPU, RAM, or storage) of an existing virtual machine or instance.
2) Horizontal Scaling: This involves adding or removing virtual machines or instances to handle fluctuating workloads.

Auto scaling: In cloud computing, Auto - Scaling (AS) is that allow user to automatic scale cloud services, such as Virtual Machine (VM) and server capacities Up or Down, depending on defining situation. Auto - scaling automates the contraction of system capacity that is available for applications and is a desired feature in cloud PaaS and IaaS

offerings. When feasible, technology buyers should use it to match provisioned capacity to application demand and reduce costs. In Amazon Web Service (AWS), auto - scaling is defined as a cloud computing service feature that allows AWS users to automatically launch or terminate virtual instances based on defined policies, health status checks, and schedules. Auto - scaling is the capability in cloud computing infrastructures that enables dynamic provisioning of virtualized resources. Resources used by cloud based applications can be automatically maximized or minimized, in that way adapting resource usage to the application requirements

Growth is a basic goal of any business. But infrastructural support is one primary necessity to facilitate this growth. Hence, apart from flexibility of operation, what any business looks for is scalability. Business scalability largely depends on the scalability of the computing system supporting the business. Business applications must be able to support growing workload by processing sudden traffic escalation efficiently. Scalability is about customer satisfaction as it can process growing extent of incoming traffic with consistent performance. More customer satisfaction means more business. Studies show that a few milliseconds of delay in page load time reduce business by significant percentage. Amazon observed 1 percent decrease in retail revenue caused by 100 - millisecond delay. Analysis by Google showed 20 percent decrease in traffic due to an additional 500 - millisecond delay in page response time. Hence, scalability ultimately becomes a vital business concern. In cloud computing, businesses can take advantage of nearly infinite scalability by using the right scale to deploy their applications. The automatic scaling facility to grow during traffic demands and shrink during less workload allows businesses to focus on their core business objectives.

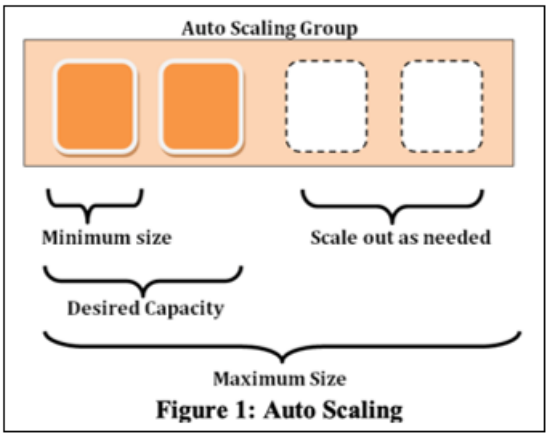

Figure 1: Auto Scaling

Importance of Scalability

Scalability is crucial for businesses for several reasons:

1) Cost Optimization: By scaling resources up or down based on demand, organizations can optimize their costs and avoid overpaying for underutilized resources.
2) Agility and Flexibility: Scalability allows businesses to quickly respond to changing market conditions, seasonal demands, or unexpected spikes in traffic.
3) High Availability: Scalable systems can distribute workloads across multiple instances, ensuring high availability and minimizing downtime.
4) Future - proofing: As businesses grow, scalable cloud solutions can accommodate increasing demands without the need for costly infrastructure upgrades. (Bhowmik, 2017)

Proposed Architecture: Scalable Web Application

In my work as a software developer, I have implemented scalable architectures for web applications using cloud computing services. One such architecture follows a microservices pattern, leveraging containerization and orchestration tools like Docker and Kubernetes.

Ethical Issues in Cloud Computing

Cloud computing implicates several ethical obligations. This is primarily caused by the fact that organizations work outside their trusted network boundary in cloud environment. Risk arises as controls are released to third party cloud vendors. The blurring network boundary raises confusion regarding accountability of participating parties. The complicated structure of cloud may sometimes raise concern about the responsibility when some problems arise. Both the cloud computing vendors and the users must have clear ideas about their individual responsibilities. There are many instances where both parties need to work together to resolve issues. Another important part of the ethical side falls on the cloud vendor's shoulders. This issue surfaces as cloud computing allows easy distribution of intellectual properties of other people. In cloud computing, users or enterprises have to rely their data upon cloud storages with the faith that security and privacy of information will be maintained by the cloud vendors. Cloud computing companies have total control over those data stored in their data centers. How they use those sensitive data depends on their moral. Any trusted cloud vendor must implement mechanism in order to provide unbreakable protection to its users' data. (Bhowmik, 2017)

Role of API

(Application Program Interface) is a set of defined functions or methods which is used to compile the application. It defines the contract of communication or standard interface provided by software components for others (other software components) in order to interact with them. APIs play important role in cloud computing. When some cloud services are released, corresponding APIs (referred as cloud API) are also released as they are critical for the usefulness and operational success of those services. Cloud services generally provide well - defined APIs for its consumers so that anyone can access and use the capabilities offered to develop application or service. Request for data or computation can be made to cloud services through cloud APIs. Cloud APIs expose their features via REST or SOAP. For instances, the cloud APIs can be classified as IaaS (Infrastructure as a Service) API for resource configuration or workload management, PaaS (Platform as a Service) API to integrate with database service etc. and SaaS (Software as a Service) API to integrate with application services like ERP or CRM. Both vendor specific and cross platform APIs are available, but cross platform APIs are still not widespread in cloud computing arena and are available for specific functional areas only. APIs streamline the access of cloud services and enforce the adherence to compliance. (Bhowmik, 2017)

Confusion between Cloud and Internet

Many people confuse Internet as cloud computing. Where however, the fact is that these two are different. The confusion arises because cloud computing is generally delivered via Internet. The actual fact is that earlier users used to access websites or web portals consisting of static and dynamic pages via Internet but now they access cloud computing too. Cloud computing has its own characteristics. It follows utility service model and it is measurable where users can be billed as per use. More importantly, it can deliver both software and hardware to users over internetwork or Internet in special form called 'service'. Cloud computing does not mean simple static or dynamic web content; it is much more than that. (Bhowmik, 2017)

Architecture Overview

1) Load Balancer: A load balancer distributes incoming traffic across multiple application instances, ensuring high availability and scalability.
2) Microservices: The application is divided into smaller, independent microservices, each responsible for a specific functionality (e. g., user authentication, product catalog, order processing).
3) Containerization: Each microservice is packaged into a Docker container, ensuring consistent deployment and portability across environments.
4) Kubernetes Cluster: The containers are deployed and managed within a Kubernetes cluster, which handles automatic scaling, load balancing, and self - healing capabilities.
5) Persistent Storage: Stateful data, such as databases or file storage, is decoupled from the application instances and managed separately, ensuring data persistence and scalability.
6) Monitoring and Logging: Comprehensive monitoring and logging solutions are implemented to track application performance, identify bottlenecks, and facilitate debugging and troubleshooting.

Scaling Strategies

1) Horizontal Scaling: Kubernetes automatically scales the number of replicas (instances) of each microservice based on predefined metrics, such as CPU utilization or request rates.
2) Vertical Scaling: If necessary, individual instances can be vertically scaled by increasing or decreasing their allocated resources (CPU, RAM, etc.).
3) Auto scaling: Auto scaling policies can be configured to automatically scale resources based on predefined thresholds, ensuring optimal resource utilization and cost - effectiveness.
4) Utility Computing: Computing as Measured and Utility Service The computing era reached a point where it was empowered with the following characteristics:
   - Scalable computing infrastructure made of heterogeneous resources that can be grown as much as required in real time
   - Single distributed computing environment spread across the globe, empowered by high speed network
   - Collaborative work facility from different locations, empowered by modern age web service standards
   - Flexible application architecture that is easily modifiable with changing business requirements, empowered by the SOA paradigm All these features combined with resource virtualization technique created scope for a new avenue of computing. Under this new model, the vendors could arrange all required computing facilities which users could consume on payment basis. Since, payment is calculated on use basis, and users need not take responsibility of system procurement and management related activities, that provide huge benefit from users' point of view. This model of computing is known as utility computing. The idea for such a model was first presented by John McCarthy, a professor at Massachusetts Institute of Technology (MIT) in 1961 which showed that the computing can be delivered as utility service much like electricity.

Enhancing Scalability with AI and ML

Artificial Intelligence (AI) and Machine Learning (ML) can play a significant role in enhancing the scalability of cloud computing solutions. Here are a few potential applications:

1) Predictive Autoscaling: ML models can analyse historical usage patterns, application logs, and other relevant data to predict future resource demands accurately. This information can be used to proactively scale resources, ensuring optimal performance and cost - effectiveness.
2) Intelligent Load Balancing: AI algorithms can dynamically adjust load balancing strategies based on real - time traffic patterns, application behaviour, and resource utilization, optimizing resource allocation and minimizing latency.
3) Anomaly Detection: ML models can detect anomalies in application performance, resource utilization, or user behaviour, enabling proactive scaling or remediation actions to maintain system stability and availability.
4) Resource Optimization: AI techniques can analyse resource usage patterns and identify opportunities for optimizing resource allocation, reducing waste, and minimizing costs.
5) Self - healing Systems: ML - powered monitoring and analysis can enable self - healing systems that can automatically detect and resolve issues, minimizing downtime and ensuring high availability.

Economic Benefits of Cloud Computing

Cloud computing has the potential to be one of the most transformative economic innovations of the twenty - first century—allowing companies and government agencies to scale resources quickly, increase portability and accessibility, reduce costs, and increase productivity. Cloud computing can transform important supply chain - reliant sectors like advanced manufacturing, chemicals, and retail. This is particularly true for SMEs, which can take advantage of network effects and lowered barriers of access for sophisticated, IT - intensive applications. The marginal Rate of return is higher for SMEs whose fixed costs can be lowered—jolted by the economies of scale of the collective resources across companies' user bases—and employ BYOD ("bring your own device") policies that reduce overhead, increase efficiency, and allow workers greater flexibility to work remotely. The US experience with cloud computing is illustrative. The United States' broadly uniform legal context, large market size, and customer base that is at ease with

digital applications primed the country to be a first mover in the cloud computing space. Of new US businesses, 69 percent attribute part of their productivity growth to cloud computing, 60 percent say cloud computing saves time, and 40 percent say it saves money.8 in the EU, the impression is similar. The European Commission estimates that every €1 spent on software as a service (SaaS) replaces €2.30 spent on traditional administrative solutions.9 Finance decision - makers in European companies believe that cloud computing brings many business benefits, such as increased flexibility (57 percent), capacity (56 percent), and scalability (53 percent). An overwhelming 96 percent believe that cloud computing provides their business with quantifiable benefits, such as reduced information technology (IT) maintenance costs, reduced IT Spending, reduced operational costs, and improved process efficiency. And 55 percent believe that cloud computing offers better value than traditional outsourcing.10The macroeconomic impact is eye - opening. The growth in cloud computing is coupled with massive new labour force demands. According to a Analytics study, there are over eighteen million cloud - dependent jobs globally with 40.8 percent of jobs in China, 21.7 percent in the United States, and 12.2 percent in India.

Extended Business Use Case

How Recommender Systems Enhance Scalability in Cloud Computing

Recommender systems, crucial for personalization in digital businesses, can significantly enhance the scalability of cloud computing environments through several mechanisms:

1) **Efficient Resource Utilization**: By predicting user preferences accurately, recommender systems can help optimize the use of computational resources in the cloud. Efficient candidate generation and ranking reduce the need for processing large datasets repeatedly, thus saving computational power and reducing costs.
2) **Dynamic Scaling**: The ability to dynamically adjust recommendations based on real - time data means cloud resources can be scaled up or down based on current demands. This dynamic scaling helps manage the load efficiently, ensuring that cloud resources are used optimally and can handle peak loads without wastage during off - peak times.
3) **Handling Large Datasets**: Recommender systems are designed to handle large amounts of data by narrowing down the recall size from thousands to hundreds of potential recommendations. This ability to filter and rank data efficiently translates into better performance and faster processing times in cloud environments, supporting scalability.
4) **Use of Implicit Feedback**: By leveraging implicit feedback, such as click - through rates and viewing times, recommender systems can improve their accuracy without needing large amounts of explicit feedback. This reduces the data processing burden on cloud systems, making it easier to scale services as the user base grows.
5) **Overcoming Challenges**: The challenges faced by recommender systems, such as cold start problems and data sparsity, can be mitigated with scalable cloud solutions. For instance, hybrid approaches that combine user - item and item - item methods, as well as using heuristics for business rules, help in creating robust systems that can scale efficiently.
6) **Advanced Algorithms**: The use of sophisticated machine learning algorithms, including neural networks and collaborative filtering, allows recommender systems to handle high - dimensional data and complex user interactions. These algorithms are highly scalable and can be deployed across distributed cloud architectures to manage large - scale data processing.

By integrating recommender systems into cloud computing environments, businesses can achieve enhanced scalability, ensuring that their services remain efficient, responsive, and capable of handling increasing amounts of data and user interactions.

2. Conclusion

Scalability is a critical aspect of cloud computing, enabling businesses to adapt to changing demands and optimize resource utilization. By leveraging scalable architectures, such as micro services and containerization, organizations can achieve high availability, agility, and cost - effectiveness. Additionally, the integration of AI and ML techniques can further enhance scalability by enabling predictive autoscaling, intelligent load balancing, anomaly detection, and resource optimization. As businesses continue to embrace cloud computing, scalability will remain a key consideration for ensuring efficient and reliable operations. he Security of the Cloud As previously mentioned, it is a misnomer to speak of the cloud as a single, cohesive entity. Therefore, the security of the cloud is much like security of the internet writ large—a broad swath of Various potential threats, vulnerabilities, and risks affecting thousands, if not millions, of different types of services provided to actors ranging from single individuals to entire governments. There will never be an incident where the cloud goes down, so to speak, as if it could be switched off like a lamp. Instead, security issues in the cloud cover a spectrum ranging from complete failure and unavailability to limited performance or effects limited to subsets of data and services. Cloud computing vendors build data centers at locations of their convenience, both geographical and economical. A vendor may even have more than one data centers dispersed over multiple geographic locations. Since subscribers remotely access cloud computing over the Internet, they may not be aware of the actual location of the resources they consume. More importantly, the storage location of subscriber’s data may not be within the country or region of the subscriber. This sometimes poses serious legal concerns. The privacy or compliance rule generally differs across different legal jurisdictions. The rules for degree of disclosure of personal data to government agencies (in cases of some official investigations) differ from country to country, or even state to state within a country. Situation may arise where the law of the country of a cloud subscriber asks for some data to be disclosed where the law of hosting region of the cloud (that is the region/country of cloud data center) does not allow such disclosure. Most regulatory frameworks recognize cloud consumer organizations responsible for the security, integrity, and storage of data even when in reality it is held by an external cloud vendors. In such scenario, resolving the multi - regional compliance and legal issues are soaring challenges before cloud computing It can be said that

cloud computing is available anywhere and anytime. But it is true that cloud computing vendors cannot generally deliver services with hundred percent service uptime assurance. Reputed cloud vendors guarantee more than 99.9% of service uptime, but till now it has not been possible for vendors to deliver 100% service uptime over a reasonable period of time. Anyway, despite of this, its service availability is still much better than the traditional computing scenario. And cloud users rarely suffer or can even feel the effect of the almost insignificant service downtime.

References

[1] V. Dastjerdi and R. Buyya, "An Autonomous Reliability - Aware Negotiating Strategy for Cloud Computing Environments, " in Proc. IEEE/ACM Int. Symp. Cluster, Cloud and Grid Computing, 2012, pp.284 - 291.
[2] L. Wu and R. Buyya, "Service Level Agreement (SLA) in Utility Computing Systems, " Information Science Reference, IGI Global, 2012, pp.1 - 25.
[3] M. Armbrust, A. Fox, R. Griffith, et al., "Above the Clouds: A Berkeley View of Cloud Computing, " Electrical Engineering and Computer Sciences (EECS), UC Berkeley, 2009, pp.1 - 23.
[4] L. M. Vaquero, L. Rodero - Merino, J. Caceres, and M. Lindner, "Dynamically Scaling Applications in the Cloud, " ACM SIGCOMM Computer Communication Review, vol.41, pp.45 - 52, 2011.
[5] P. Rimal, E. Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems, " in Proc. IEEE Int. Joint Conf. INC, IMS and IDC, 2009, pp.44 - 51.
[6] P. Covington, J. K. Adams, and E. Sargin, "Deep Neural Networks for YouTube Recommendations, " in *Proc.10th ACM Conf. Recommender Systems*, 2016, pp.191 - 198.
[7] Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review, " *Expert Syst. Appl.*, vol.97, pp.205 - 227, May 2018.
[8] R. J. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, and J. D. Knowles, "Fifty years of pulsar candidate selection: from simple filters to a new principled real - time classification approach, " *Mon. Not. R. Astron. Soc.*, vol.459, no.1, pp.1104 - 1123, May 2016.
[9] Yogananda Domlur Seetharama, Enhancing Personalization Through the Impact of Recommender Systems, International Journal of Advanced Research in Engineering and Technology (IJARET), 9 (6), 2018, pp 298 - 315. https: //iaeme. com/Home/issue/IJARET?Volume=9&Issue=6