Cloud Computing : An Overview

Shubham Prasad[1] , Gaurav Kumar Singh[2] , Mr Abhishek Kumar Singh[3]
[1,2,3] Department Of MCA, IIMT College of Engineering, Greater Noida

Abstract - Cloud computing has emerged as a transformative paradigm in information technology, offering scalable, on-demand access to computing resources over the internet. This paper provides a comprehensive overview of cloud computing, covering its fundamental concepts, service models (IaaS, PaaS, SaaS), and deployment models (public, private, hybrid, and community clouds). The study explores the key benefits of cloud computing, including cost efficiency, flexibility, scalability, and enhanced collaboration, while also addressing major challenges such as data security, privacy concerns, and vendor lock-in. Additionally, the paper discusses current trends and future directions in cloud technology, including the integration of artificial intelligence, edge computing, and containerization. This overview aims to serve as a foundational reference for researchers, IT professionals, and decision-makers seeking to understand the evolving landscape of cloud computing and its implications for modern computing environments.

Index Terms - IaaS, PaaS, SaaS , Virtualization

INTRODUCTION

In recent years, cloud computing has revolutionized the way individuals and organizations manage, process, and store data. As a model for delivering computing services—including servers, storage, databases, networking, software, and analytics—over the internet, cloud computing enables users to access resources on-demand with minimal management effort. This shift from traditional on-premise IT infrastructure to cloud-based systems has significantly improved operational efficiency, reduced capital expenditure, and enhanced scalability.

Cloud computing is built on a foundation of virtualization and distributed computing, allowing dynamic resource allocation and multi-tenant environments. It offers three primary service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). These services are deployed through various models, including public, private, hybrid, and community clouds, each offering different levels of control, security, and flexibility.

The adoption of cloud computing spans across industries such as healthcare, finance, education, and government, driven by the need for innovation, agility, and digital transformation. Despite its numerous advantages, cloud computing also presents challenges related to data privacy, security, compliance, and service reliability. Therefore, understanding its core principles, benefits, and limitations is essential for stakeholders seeking to leverage this technology effectively.

This paper aims to provide a comprehensive overview of cloud computing, highlighting its architecture, service and deployment models, advantages, challenges, and emerging trends shaping its future.

HISTORY AND STATUS

With the explosion of the Internet, tight pressure is put to the existing storage and computing facilities. The Internet service providers start to use the cheap commodity PCs as the underlying hardware platform. Various kinds of software technologies are invented to make these PCs work elastically, which has led to 3 major cloud computing styles based on the underlying resource abstraction technologies: the Amazon style, Google Style and Microsoft style.

 Amazon's cloud computing is based on server virtualization technology. Amazon released Xen-based Elastic Compute Cloud™ (EC2), object storage service (S3) and structure data storage service (SimpleDB)[12] during the 2006 – 2007, under the name Amazon Web Service™ (AWS)[9]. Ondemand and cheaper AWS becomes the pioneer of Infrastructure as a Service (IaaS) provider.

 Google's style is based on technique-specific sandbox. Google published several research papers from 2003 to 2006[1-5], which outline a kind of Platform as a Service (PaaS) cloud computing. The platform, which is called Google App Engine™ (GAE), is released to public as a service in 2008.

 Microsoft Azure™ [10] is released in Oct. 2008, which uses Windows Azure Hypervisor (WAH) as the underlying cloud infrastructure and .NET as the application container. Azure also offers services including BLOB object storage and SQL service.

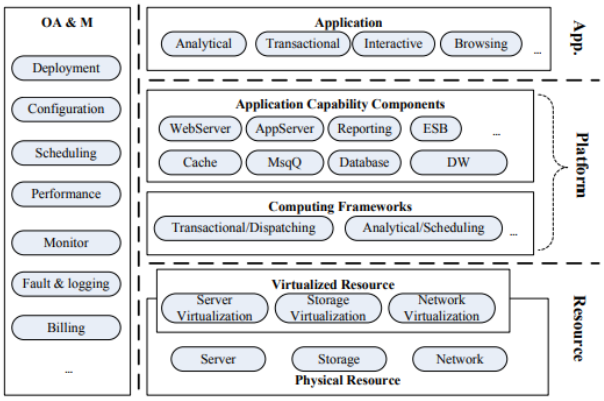
It's hard to judge which one is better, but apparently server virtualization is more flexible and compatible with existing software and applications; while the sandboxes put more restrictions on programming languages but less abstraction .

CLOUD COMPUTING ARCHITECTURE

Several organizations and researchers have provided the architecture of cloud computing. In essence, the entire system is split into the core stack and the management. Within the core stack, three layers exist: (1) Resource (2) Platform and (3) Application. The resource layer is the infrastructure layer that consists of physical and virtualized computing, storage and network resources.

The platform layer is the most complicated component that may be split into numerous sublayers. For example, a computing framework handles the transaction dispatching and/or task scheduling. A storage sub-layer offers unlimited storage and caching ability.

The application server and other parts enable the same general application logic as previously with either on-demand capability or flexible management, so that no components will be the bottle neck of the entire system**.**



Based on the underlying resource and components, the application could support large and distributed transactions and management of huge volume of data. All the layers provide external service through web service or other open interfaces.

CLOUD COMPUTING CATEGORIES

There are various dimensions to categorize cloud computing, two of the most widely used categories are: service boundary and service type.

* From the perspective of service boundary, cloud computing can be categorized as public
cloud, private cloud and hybrid cloud. The public cloud is services offered to
outside parties. The companies create and manage private cloud on their own. Hybrid cloud is the sharing of resources of public cloud and private cloud through a secure network. Virtual Private Cloud (VPC) services launched by Google[8] and Amazon[9] are some instances of Hybrid cloud.

* From the perspective of the type of service, cloud computing can be categorized as Infrastructure
as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS).

ADVANTAGES AND RISKS

The cloud computing is a Win-Win strategy for the service provider and the service consumer. We summarize the advantages as below:

Satisfy business requirements on demand by resizing the resource occupied by application to fulfill the changing the customer requirements.

Lower cost and energy-saving. By making use of low cost PC, customerized low power consuming hardware and server virtualization, both CAPEX and OPEX are decreased.

Improve the efficiency of resource management through dynamic resource scheduling. However there are also some major challenges to be studied.

Privacy and security. Customer has concerns on their privacy and data security than traditional hosting service.

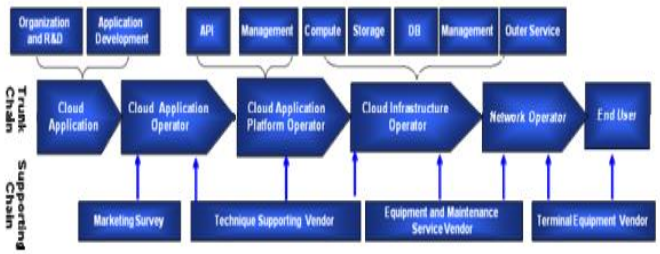
The continuity of service. It refers to the factors that may negatively affected the continuity of cloud computing such as Internet problems, power cut-off, service disruption and system bugs. Followed are some typical cases of such problems: In November 2007, RackSpace, Amazon's competitor, stopped its service for 3 hours because of power cut-off at its data center; in June 2008, Google App Engine service broke off for 6 hours due to some bugs of storage system; In March 2009, Microsoft Azure experienced 22 hours' out of service caused by OS system update. Currently, the public cloud provider based on virtualization, defines the reliability of service as 99.9% in SLA.

Service migration. Currently, no regularity organization have reached the agreement on the standardization of cloud computing's external interface.

As a result, once a customer started to use the service of a cloud computing provider, he is most likely to be locked by the provider, which lay the customer in unfavorable conditions.

VALUE CHAIN OF CLOUD COMPUTING

The following figure depicts the cloud computing value chain with related organizations and their functionalities.



Cloud Applications: The driven force of cloud computing, which is different from traditional application module. Cloud Application Operator: Offers cloud computing products. In many cases, they are same as by application provider or platform provider.

Cloud Application Platform Operator: Provides cloud application and development platform, such as GAE™ and Force.com™, etc. Cloud infrastructure operator: Provide infrastructure service, such as AWS™ and GoGrid.

Network Operator: Provide network access service to the above platform operators and the end users.

Technique supporting vendor: Offer technical support to players in this chain, including software development, testing,

provisioning and operation.

Terminal equipment vendor: Offer device maintenance service for all players in the chain.

End Users: End users pays for the cloud services.

STANDARDIZATION

Distributed Management Task Force[7] is an industry alliance composed by over 200 IT related corporations including IBM, EMC, HP, Cisco, Oracle and Microsoft, which is committed to develop, maintain and popularize the IT management system under enterprise context. DMTF has published Virtualization Management Initiative and Open Virtualization Format, and the latter is declared to be supported by major vendors. DMFT founded Open Cloud Standards Incubator at the year 2009, whose aim is to clarify the interoperability between several cloud systems.

Distributed Management Task Force[7] is an industry alliance composed by over 200 IT related corporations including IBM, EMC, HP, Cisco, Oracle and Microsoft, which is committed to develop, maintain and popularize the IT management system under enterprise context. DMTF has published Virtualization Management Initiative and Open Virtualization Format, and the latter is declared to be supported by major vendors. DMFT founded Open Cloud Standards Incubator at the year 2009, whose aim is to clarify the interoperability between several cloud systems.

Besides, SNIA (Storage Network Industry Association), CSA (Cloud Security Alliance) and OCC (Open Cloud Consortium) is now working on cloud storage, cloud security and cloud intercommunication standards respectively. In order to coordinate the work of above standardization organizations, OMG (Object Management Group) appealed that all the organizations maintain their own standards on http://cloudstandards.org.

PaaS and SaaS don't have related cloud computing standards yet. Most current systems exploit mature protocols and have variety kinds of service forms.

CLOUD INTEROPERABILITY

Cloud interoperability refers to customers' ability to use the same artifacts, such as management tools, virtual server images, and so on, with a variety of cloud computing providers and platforms.

Cloud interoperability will enable cloud infrastructures to evolve into a worldwide, transparent platform in which applications aren't restricted to enterprise clouds and cloud service providers. We must build new standards and interfaces that will enable enhanced portability and lexibility of virtualized applications. Up to now, signiicant discussion has occurred around open standards for cloud computing. In this context, the "Open Cloud Manifesto" (www.open cloudmanifesto.org) provides a minimal set of principles that will form a basis for initial agreements as the cloud community develops standards for this new computing paradigm.

SECURITY AND PRIVACY

In cloud computing, a data center holds information that end-users would more traditionally have stored on their computers. This raises concerns regarding user privacy protection because users must outsource their data. Additionally, the move to centralized services could affect the privacy and security of users' interactions. Security threats might happen in resource provisioning and during distributed application execution. Also, new threats are likely to emerge. For instance, hackers can use the virtualized infrastructure as a launching pad for new attacks. Cloud services should preserve data integrity and user privacy. At the same time, they should enhance interoperability across multiple cloud service providers. In this context, we must investigate new dataprotection mechanisms to secure data privacy, resource security, and content copyrights.

KEY CHALLENGES

In 1961, John McCarthy envisioned that "computation may someday be organized as a public utility." We can view the cloud computing paradigm as a big step toward this dream. To realize it fully, however, we must address several signiicant problems and unexploited opportunities concerning the deployment, eficient operation, and use of cloud computing infrastructures.

Software/Hardware Architecture

Cloud computing services's emergence suggests fundamental changes in software and hardware architecture. Computer architectures should shift the focus of Moore's law from increasing clock speed per chip to increasing the number of processor cores and threads per chip. Industry and academia must design novel systems and services that would exploit a high degree of parallelism. Software architectures for massively parallel, data-intensive computing, such as MapReduce (http://labs.google.com/papers/ mapreduce.html), will grow in popularity. In terms of storage technologies, we'll likely shift from hard disk drives (HDDs) to solid-state drives (SDDs), such as lash memories, or, given that completely replacing hard disks is prohibitively expensive, hybrid hard disks — that is, hard disks augmented with lash memories, which provide reliable and high-performance data storage. The biggest barriers to adopting SSDs in data centers have been price, capacity, and, to some extent, the lack of sophisticated query-processing techniques. However, this is about to change as SSDs' I/O operations per second (IOPS) beneits become too impressive to ignore, their capacity increases at a fast pace, and we devise new algorithms and data structures tailored to them.

DATA MANAGEMENT

The shift of computer processing, storage, and software delivery away from desktop and local servers, across the Internet, and into nextgeneration data centers results in limitations as well as new opportunities regarding data management. Data is replicated across large geographic distances, where its availability and durability are paramount for cloud service providers. It's also stored at untrusted hosts, which creates enormous risks for data privacy. Computing power in clouds must be elastic to face changing conditions. For instance, providers can allocate additional computational resources on the ly to handle increased demand. They should deploy novel data management

approaches, such as analytical data management tasks, multitenant databases for SaaS, or hybrid designs among database management systems (DBMSs) and MapReduce-like systems so as to address data limitations and harness cloud computing platforms' capabilities.

SERVICE PROVISIONING AND CLOUD ECONOMICS

Providers supply cloud services by signing service-level agreements (SLAs) with consumers and end-users. Cloud service consumers, for instance, might have an SLA with a cloud service provider concerning how much bandwidth, CPU, and memory the consumer can use at any given time throughout the day. Underestimating the provision of resources would lead to broken SLAs and penalties. On the other hand, overestimating the provision of resources would lead to resource underutilization and, consequently, a decrease in revenue for the provider. Deploying an autonomous system to eficiently provision services in a cloud infrastructure is a challenging problem due to the unpredictability of consumer demand, software and hardware failures, heterogeneity of services, power management, and conlicting signed SLAs between consumers and service providers.

In terms of cloud economics, the provider should offer resource-economic services. Novel, power-eficient schemes for caching, query processing, and thermal management are mandatory due to the increasing amount of waste heat that data centers dissipate for Internetbased application services. Moreover, new pricing models based on the pay-as-you-go policy are necessary to address the highly variable demand for cloud resources.

In this Issue

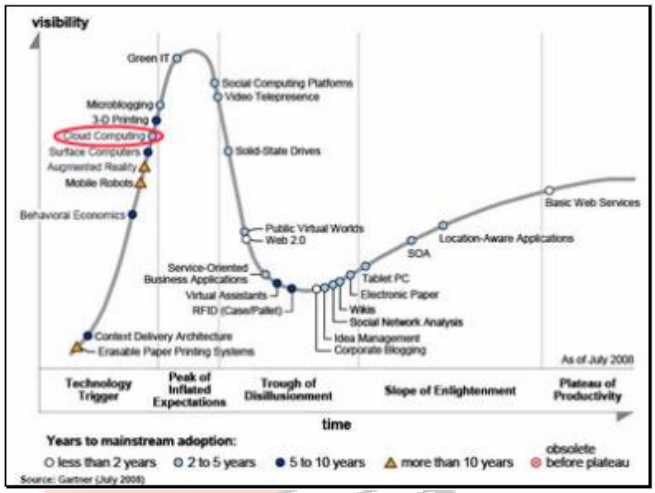
Given the continued, intense activity in the cloud arena, we invited researchers and practitioners to submit articles to this special issue of IC describing research efforts and experiences concerning the deployment, eficient operation, and use of cloud computing infrastructures. From among the 42 submissions, and after rigorous review, we selected the following four articles as representative of ongoing research and development activities.

The irst article, "Virtual Infrastructure Management in Private and Hybrid Clouds," by Borja Sotomayor, Rubén S. Montero, Ignacio M. Llorente, and Ian Foster, presents two open source projects for private and hybrid clouds. OpenNebula is a virtual infrastructure manager that can be used to deploy virtualized services on both a local pool of resources and on external IaaS clouds. Haizea is a resource lease manager that can act as a scheduling back end for OpenNebula, providing advance reservations and resource preemption.

"Harnessing Cloud Technologies for a Virtualized Distributed Computing Infrastructure," by Alexandre di Costanzo, Marcos Dias de Assunção, and Rajkumar Buyya, presents the realization of a system — termed the InterGrid — for interconnecting distributed computing infrastructures by harnessing virtual machines. The article provides an abstract view of the proposed architecture and its implementation. Experiments show the scalability of an InterGridmanaged infrastructure and how the system can beneit from using cloud infrastructure.

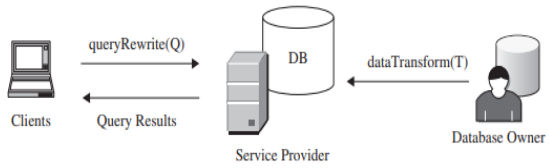
In "Content-Centered Collaboration Spaces in the Cloud," John S. Erickson, Susan Spence, Michael Rhodes, David Banks, James Rutherford, Edwin Simpson, Guillaume Belrose, and Russell Perry envision a cloud-based platform that inverts the traditional application-content relationship by placing content rather than applications at the center, letting users rapidly build customized solutions around their content items. The authors review the dominant trends in computing that motivate the exploration of new approaches for content-centered collaboration and offer insights into how certain core problems for users and organizations are being addressed today.

The inal article, "Sky Computing," by Katarzyna Keahey, Maurício Tsugawa, Andréa Matsunaga, and José A.B. Fortes, describes the creation of environments conigured on resources provisioned across multiple distributed IaaS clouds. This technology is called sky computing. The authors provide a realworld example and illustrate its beneits with a deployment in three distinct clouds of a bioinformatics application.



DATABASE OUTSOURCING AND QUERY INTEGRITY ASSURANCE

In recent years, database outsourcing has become an important component of cloud computing. Due to the rapid advancements in network technology, the cost of transmitting a terabyte of data over long distances has decreased significantly in the past decade. In addition, the total cost of data management is five to ten times higher than the initial acquisition costs. As a result, there is a growing interest in outsourcing database management tasks to third parties that can provide these tasks for a much lower cost due to the economy of scale. This new outsourcing model has the benefits of reducing the costs for running Database Management Systems (DBMS) independently and enabling enterprises to concentrate on their main businesses [12]. Figure 8.7 demonstrates the general architecture of a database outsourcing environment with clients. The database owner outsources its data management tasks, and clients send queries to the untrusted service provider. Let T denote the data to be outsourced. The data T are is preprocessed, encrypted, and stored at the service provider. For evaluating queries, a user rewrites a set of queries Q against T to queries against the encrypted database. The outsourcing of databases to a third-party service provider was first introduced by Hacigu¨mu¨s et al. [13]. Generally, there are two security concerns in database outsourcing. These are data privacy and query integrity. The related research is outlined below.



Data Privacy Protection. Hacigu¨mu¨s et al. [37] proposed a method to execute SQL queries over encrypted databases. Their strategy is to process as much of a query as possible by the service providers, without having to decrypt the data. Decryption and the remainder of the query processing are performed at the client side. Agrawal et al. [14] proposed an order-preserving encryption scheme for numeric values that allows any comparison operation to be directly applied on encrypted data. Their technique is able to handle updates, and new values can be added without requiring changes in the encryption of other values. Generally, existing methods enable direct execution of encrypted queries on encrypted datasets and allow users to ask identity queries over data of different encryptions. The ultimate goal of this research direction is to make queries in encrypted databases as efficient as possible while preventing adversaries from learning any useful knowledge about the data. However, researches in this field did not consider the problem of query integrity.

Query Integrity Assurance. In addition to data privacy, an important security concern in the database outsourcing paradigm is query integrity. Query integrity examines the trustworthiness of the hosting environment. When a client receives a query result from the service provider, it wants to be assured that the result is both correct and complete, where correct means that the result must originate in the owner's data and not has been tampered with, and complete means that the result includes all records satisfying the query. Devanbu et al. [15] authenticate data records using the Merkle hash tree [16], which is based on the idea of using a signature on the root of the Merkle hash tree to generate a proof of correctness. Mykletun et al. [17] studied and compared several signature methods that can be utilized in data authentication, and they identified the problem of completeness but did not provide a solution. Pang et al. [18] utilized an aggregated signature to sign each record with the information from neighboring records by assuming that all the records are sorted with a certain order. The method ensures the completeness of a selection query by checking the aggregated signature. But it has difficulties in handling multipoint selection query of which the result tuples occupy a noncontinuous region of the ordered sequence. The work in Li et al. [19] utilizes Merkle hash tree-based methods to audit the completeness of query results, but since the Merkle hash tree also applies the signature of the root Merkle tree node, a similar difficulty exists. Besides, the network and CPU overhead on the client side can be prohibitively high for some types of queries. In some extreme cases, the overhead could be as high as processing these queries locally, which can undermine the benefits of database outsourcing. Sion [20] proposed a mechanism called the challenge token and uses it as a probabilistic proof that the server has executed the query over the entire database. It can handle arbitrary types of queries including joins and does not assume that the underlying data is ordered. However, the approach is not applied to the adversary model where an adversary can first compute the complete query result and then delete the tuples specifically corresponding to the challenge tokens [21]. Besides, all the aforementioned methods must modify the DBMS kernel in order to provide proof of integrity.

Recently,Wang et al. [22] proposed a solution named dual encryption to ensure query integrity without requiring the database engine to perform any special function beyond query processing. Dual encryption enables cross-examination of the outsourced data, which consist of (a) the original data stored under a certain encryption scheme and (b) another small percentage of the original data stored under a different encryption scheme.
Users generate queries against the additional piece of data and analyze their results to obtain integrity assurance. For auditing spatial queries, Yang et al [23] proposed the MR-tree, which is an authenticated data structure suitable for verifying queries executed on outsourced spatial databases. The authors also designed a caching technique to reduce the information sent to the client for verification purposes. Four spatial transformation mechanisms are presented in Yiu et al. [24] for protecting the privacy of outsourced private spatial data. The data owner selects transformation keys that are shared with trusted clients, and it is infeasible to reconstruct the exact original data points from the transformed points without the key. However, both aforementioned researches did not consider data privacy protection and query integrity auditing jointly in their design.

The state-of-the-art technique that can ensure both privacy and integrity for outsourced spatial data is proposed in Ku et al. [12]. In particular, the solution first employs a one-way spatial transformation method based on Hilbert curves, which encrypts the spatial data before outsourcing and hence ensures its privacy. Next, by probabilistically replicating a portion of the data and encrypting it with a different encryption key, the authors devise a mechanism for the client to audit the trustworthiness of the query results.

While the transparent cloud provides flexible utility of network-based resources, the fear of loss of control on their data is one of the major concerns that prevent end users from migrating to cloud storage services. Actually it is a potential risk that the storage infrastructure providers become self-interested, untrustworthy, or even malicious. There are different motivations whereby a storage service provider could become untrustworthy—for instance, to cover the consequence of a mistake in operation, or deny the vulnerability in the system after the data have been stolen by an adversary. This section introduces two technologies to enable data owners to verify the data integrity while the files are stored in the remote untrustworthy storage services.

REFERENCES

[1] Ghemawat, S., Gobioff, H., Leung, S.-T.: The Google File System. In: SOSP (2003)
[2] Dean, J., Ghemawat, S.: MapReduce: Simplifed Data Processing on Large Clusters. In: OSDI 2004 (2004)
[3] Chang, F., Dean, J., Ghemawat, S., et al.: Bigtable: A Distributed Storage System for Structured Data. In: OSDI 2006 (2006)
[4] Burrows, M.: The Chubby lock service for loosely-coupled distributed systems. In: OSDI 2006 (2006)
[5] Pike, R., Dorward, S., Griesemer, R., Quinlan, S.: Interpreting the Data: Parallel Analysis with Sawzall. Scientific Programming (2005)
[6] Open Cloud Computing Interface, http://www.occi-wg.org/doku.php
[7] Distributed Management Task Force, http://www.dmtf.org
[8] Google App Engine, http://appengine.google.com
[9] Amazon Web Service, http://aws.amazon.com
[10] Microsoft Azure, http://www.microsoft.com/azure/
[11] DeCandia, G., Hastorun, D., Jampani, M., et al.: Dynamo: Amazon's Highly Available Key-value Store. In: SOSP 2007 (October 2007)
[12] Schmidt, E.: Conversation with Eric Schmidt hosted by Danny Sullivan. In: Search Engine Strategies Conference (August 2006)