Cloud Computing: An Overview

Adam Camerer
Missouri University of Science and Technology
Department of Computer Science
Rolla, MO
ajc3xc@umsystem.edu

Yannis Fu
Missouri University of Science and Technology
Department of Computer Science
Rolla, MO
ybf62n@umsystem.edu

***Abstract*—This paper gives an overview of cloud computing, giving a background its history, functionality, and applications. It covers how technologies like virtualization, distributed systems, and the internet have all combined to create the current incarnation of cloud computing, and the usefulness of the cloud ecosystem to solve the business, industrial, and scientific problems of the 21st century.**

Keywords—cloud computing; cloud operating systems; virtualization; distributed systems; fault tolerance;

I. An Introduction to Cloud Computing

In basic terms, cloud computing is the process of sharing information and resources over the internet. Cloud computing is run on cloud operating systems, which are specialized in virtualization and other tasks that are out of the scope of a regular operating system (OS) [7]. Compared to a regular OS, cloud operating systems operate on a greater scale and run on lots of computing infrastructure, connecting hardware from clusters of machines into one virtual server [5]. These servers can store large amounts of data, files, and applications on its distributed filing systems [6]. To compensate, high levels of abstraction like APIs are used to make a comfortable user experience [4].

According to the National Institute of Standards and Technology (NIST), cloud computing strives to fulfill the following characteristics:

1. On-Demand Self Computing: Where server processing time and storage spaces are allocated without the need of user input [4].
2. Broad Network Access: Allowing a wide range of thin or thick clients to access the cloud at any time without compromise [4].
3. Resource Pooling: When cloud providers to pool resources together to serve multiple customers and each of their specific needs [4].
4. Rapid Elasticity: The Ability of the cloud provider to make their services scalable based on the needs of the customer, making the architecture flexible [4].
5. Measured Service: Cloud providers monitor which services their customers use for effective resource usage and determine how they will be billed [4].

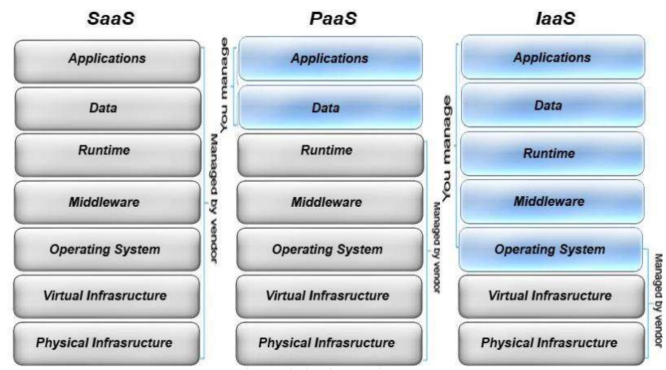

Fig. 1. The three primary service delivery types [7].

II. Cloud Models and Services

A. Service Delivery Types

There are two parts of cloud computing: the client side and the server side. The client side is a user or a customer that is accessing the cloud OS through an interface or API. The server side of cloud computing is controlled by cloud service providers (CSP) who provide services with service delivery types to the client [6]. There are three primary types of service delivery types:

1. Platform-as-a-Service (PaaS)

PaaS gives users a development environment through the internet that allows for the building and deploying of applications [4]. It is a middleware that provides databases, development tools, security, and web services to developers.

2. Infrastructure-as-a-Service (IaaS)

IaaS is a service that provides clients with IT infrastructure such as servers, storage, networks, and configured machines. These servers allow users to store and back up data in data centers at different geographical locations and provide recovery methods to restore any lost or damaged files [4].

3. Software-as-a-Service (SaaS)

SaaS hosts and runs a selection of software and applications on a cloud platform provided by the cloud provider. These applications are accessible to the user without the need to purchase or install each application. Multiple users can access the software at once and whenever they want over the internet [4]. AppsAnywhere is an example of a SaaS, where it is a platform that provides hundreds of applications for students on campus to log in and use without any other complications.

B. Cloud Development Models

There are also 4 different kinds of cloud development models that provide specific kinds of cloud environments for different purposes.

1. Public Clouds

Public clouds are catered towards being used by the general public. They are on a pay-as-you-go model that is determined by how much resources a customer consumes. Their services can be accessed on-demand through a browser or on a public network anywhere. Public clouds typically are reliable and highly scalable, but not as customizable or secure as private clouds [4].

2. Private Clouds

Private clouds are typically used in data centers owned by organizations or governments. Unlike public clouds, private clouds heavily prioritize the security of data and customization/control over the cloud infrastructure. Private clouds can be operated on or off site and are seen as efficient and price effective but with limited scalability [4].

3. Community Clouds

Community clouds are clouds managed by organizations that have mutually similar goals and interests. These systems may be run on or off site and can be outsourced to a third party to keep it running [4].

4. Hybrid Clouds

Hybrid clouds combine aspects of public, private, and community clouds into one package. Many popular cloud services like AWS and Microsoft Azure are on hybrid cloud models. Typically, the most essential services in the hybrid cloud are dedicated towards a private cloud and the lesser ones are under a public cloud [4]. The use of public clouds makes hybrids less secure than private clouds. Each cloud in the hybrid is still seen as a separate system but are connected into one platform [4].

III. History

Cloud Computing has quickly emerged in the past decade as an ecosystem to flexibly deal with disparate problems without the need for costly and complex computer infrastructure [1]. However, the buildup to this rapid explosion is decades in the making.

Initially, users shared a mainframe computer, which had time sharing to allow users to run multiple programs at once by sharing CPU time [1]. Users would connect to the mainframe through terminals, writing programs and executing them on the same computer. The main issue with it was the slowness and difficulty in maintaining the mainframe, which made it a less than optimal environment for development. Furthermore, each user had to run and execute their programs on the same OS, which made isolation impossible. Virtual machines in the 1970s allowed for multiple operating systems to be run on isolated environments, allowing for one user's work to not interfere with another [1].

With the widespread adoption of the internet in the 1990s and 2000s, along with massive advances in computer hardware, the foundations of cloud computing were created. Client-Server technologies emerged that allowed people to easily request and receive information from remote servers, making the very idea of a 'cloud' possible. With server requests and computing requirements increasing, it became cheaper to use multiple computers instead of one to keep up with the demand. However, this led to a slew of issues for coordinating and sharing data among computers [1]. Initial approaches to monitor and manage resources across a network of computers was centralized, and was prone to errors and a lack of scalability [2]. Development continued, and the first prototypes of truly distributed systems were made, with Astrolabe and SWORD using decentralized gossiping protocols to communicate among machines [2].

In the 2000s and early 2010s, companies began to offer rentable computers, storage, and online hosted software and apps [1]. Open source projects like Openstack and Hadoop helped accelerate and standardize the technology behind cloud computing. These projects provided accessible repositories for virtualizing computing, networking, and storage. This effectively increased performance, API programmability, and fault-tolerance (i.e. being able to deal with nodes or clusters failing) [3].

Containers present an opportunity for more light weight environments to run software instead of heavier virtual machines with huge overhead [3]. This would be especially important for microservices, which need a limited space to implement each features like containers that allow greater flexibility and scaling. However, they lack the full performance and security isolating features of virtual machines, making it necessary to run them on the cloud. Hardware details are being increasingly abstracted away in command line APIs and graphical websites that interface with the cloud. This means the end consumer doesn't need to know the details to develop or use cloud hosted software solutions, making the cloud more accessible. [3].

IV. Operations Breakdown

A. Virtualization

Cloud computing would not be possible without a technology known as virtualization. Virtualization creates an abstraction between the physical hardware of a datacenter and the 'logical' resources of a given application [2]. It creates an isolated environment much like a sandbox, even for processes running on the same node. This means that performance can be compartmentalized and easily determined for each process, along with having the security benefits of one process not corrupting or infecting another [2].

B. Dynamic Allocation

With processes being compartmentalized by virtualization, this allows for dynamic allocation of resources. Individual processes can be distributed among different nodes, or entire VMs can be checkpointed and migrated between nodes and clusters[2]. In some services like AWS, VMs can even be migrated while actively running. As a result, this allows for resource management to be optimized for elastic computation (i.e. quickly deal with spikes in demand for resources) and load balancing (i.e. ensuring all servers are being utilized to optimize performance), among other optimizations [2].

C. Fault Tolerance

Cloud computing is generally done in data centers, where an economy of scale allows for mass quantities of computer hardware to be purchased and utilized. Data centers are composed of individual cloud machines called nodes, which are then organized into sets of nodes known as clusters [2]. Clusters can greatly vary in size, and the nodes of a cluster are interconnected in a secluded location to prevent contamination and allow for quick communication among them [2]. Because commodity grade hardware and networking equipment is used, failures of nodes, clusters, or even entire data centers is probable and must be expected [2]. Counter-measures, colloquially known as fault-tolerance, quickly detect failures and enact failsafe measures by loading backups and switching to still functioning computing hardware to continue operations smoothly [2].

D. User Access

To interact with the cloud, users use a Command-Line-Interface API or connect to a website through a browser. Depending on the services being used, the layers of abstraction from the hardware will differ [3]. Higher level tools like web apps or a cloud drive will be much more abstracted from the hardware details than renting a virtual server, or bulk storage. Users will authenticate themselves to access the cloud resources, usually through an access key and or login credentials for the website or CLI to access cloud resources. Similar to shared personal computers, users will have different access levels and permissions.

V. Applications

The service delivery types provided by CSPs, which are SaaS, PaaS, IaaS, offer different models that serve different applications to their clients. Because of the development capabilities provided by PaaS, many users will lean towards that model when handling large databases, business intelligence, and working with app development, testing, or deployment using languages like Java and SQL [4]. Many businesses go for the SaaS route because of the IT infrastructure and systems it provides with its office software [4]. These include word editors, spreadsheet makers, presentation applications, and email systems. There are also software packages called Customer Relationship Management (CRM) dedicated for financial use, like with customer billing, invoices, payroll, and taxes [4]. IaaS services are used by businesses and enterprises as a Content Delivery Network (CDN), which provide high performance, constant backup and recovery, and are highly scalable [4].

A. Reliable Storing of Data

One of the main factors that makes cloud computing such an attractive model is its specialization in storing data. This is what gets many businesses and enterprises to start investing in cloud computing, as it offers a highly scalable storage solution that is less dependent on hardware. Physical server racks and storage devices are comparably more expensive, and are less scalable and require more on site maintenance as a result. In addition, cloud storage is much more secure, as the cloud will encrypt and require a key pair to access the data [6].

Physical storage devices such as hard drives are much more prone to failure compared to storing it in the cloud. When files are stored on the cloud, it is backed up and stored in other data centers across the world. These backups are done automatically, which are very convenient to businesses to combat a sudden system outage or failure [6]. In the case of soft/hardware failure, the cloud can recover the most important files lost on the drives. Additionally, when a system information leak occurs, the host cluster of VMs transfer the main files to a recovery system [6]. These practices provide a strong sense of fault tolerance and redundancy, minimizing the loss of data.

B. Industrial Applications

In an industrial environment, there are several benefits that an enterprise cloud operating system can provide. Enterprise cloud OSs perform incredibly large scale hardware and software management [5]. It provides several management features such as virtualization, system configuration, distributed file systems, network resources, and more in a streamlined interface [5]. Enterprise data center hardware is incorporated into a single server with tons of storage and configured network services such as a domain name system and IP addresses, [5] establishing the backbone of the IT infrastructure.

VI. Conclusion

As our lives become increasingly digitized and data from our environment and machines continues to grow at an unprecedented rate, cloud computing is needed more than ever. From storing massive amounts of data to computing meaningful information, cloud computing is quickly becoming essential for individuals, organizations, and governments. Founded on a mixture of old and new technologies, cloud computing has already revolutionized our lives in its short existence, but it is just the beginning. As the cloud computing ecosystem grows and matures, it will emerge as one of the bedrocks of the 21st century.

References

[1] Jatankumar Sedani, Minal Doshi, "Cloud Computing:From The Era Of Beginning To Present" *International Journal of Novel Research in Computer Science and Software Engineering,* Vol. 2, Issue 2, pp: (33-38), Month: May - August 2015

[2] F. Pianese, P. Bosch, A. Duminuco, N. Janssens, T. Stathopoulos and M. Steiner, "Toward a Cloud Operating System," *2010 IEEE/IFIP Network Operations and Management Symposium Workshops*, Osaka, Japan, 2010, pp. 335-342. 10.1109/NOMSW.2010.5486552.

[3] Chen, Zuo-Ning, et al. "Evolution of Cloud Operating System: From Technology to Ecosystem." *Journal of Computer Science and Technology*, vol. 32, no. 2, Mar. 2017, pp. 224–241, https://doi.org/10.1007/s11390-017-1717-z

[4] I. Odun-Ayo, M. Ananya, F. Agono and R. Goddy-Worlu, "Cloud Computing Architecture: A Critical Analysis," 2018 18th International Conference on Computational Science and Applications (ICCSA), Melbourne, VIC, Australia, 2018, pp. 1-7, doi: 10.1109/ICCSA.2018.8439638.

[5] G. Xiong, T. Ji, X. Zhang, F. Zhu and W. Liu, "Cloud operating system for industrial application," 2015 IEEE International Conference on Service Operations And Logistics, And Informatics (SOLI), Yasmine Hammamet, Tunisia, 2015, pp. 43-48, doi: 10.1109/SOLI.2015.7367408.

[6] Cao, F., Zhang, L., Naik, D. A., Gonzáles, J. L. A., Verma, N., Jain, A., Jain, R., & Sharma, A. (2022). Application of Cloud Computing Technology in Computer Secure Storage. Scientific Programming, 1–9. https://doi.org/10.1155/2022/4767725.

[7] Khajehei, K. (2014). Role of virtualization in cloud computing. International Journal of Advance Research in Computer Science and Management Studies, 2(4), 0.