UNLEASHING CLOUD POTENTIAL: A COMPREHENSIVE EXPLORATION

Santosh Venkatraman, Tennessee State University
Muhammed Miah, Tennessee State University

ABSTRACT

Traditional on-premise computing is giving way to a transformative era, the Cloud computing era, where organizations are increasingly embracing cloud-based solutions to gain a strategic edge. By outsourcing computing needs to providers like Amazon Web Services and Microsoft Azure, organizations unlock agility and refocus resources on core business and customer-centric endeavors. Cloud computing's various models allow optimal cost efficiency and computational scalability without the burden of hardware and software management. Beyond organizational benefits, cloud computing democratizes access to computing resources, as smaller organizations can empower developers to launch global apps, researchers gain unprecedented data analysis capabilities, and end-users access abundant software and storage for digital media.

The paper provides a comprehensive understanding of cloud computing, encompassing its nature, delivery and deployment models. It explores applications, assesses advantages and disadvantages, describes case studies, analyzes major vendor, associated costs, and finally speculates on future trends. In summary, this article illuminates the transformative role of cloud computing in reshaping organizational paradigms and fostering information technology accessibility across diverse organizations.

INTRODUCTION

Organizations are rapidly turning to Cloud based computing solutions, as it is becoming clear that it is a strategic advantage, and freeing up the organizational resources to focus on their core business and their customers. Cloud computing allows organizations to essentially outsource their computing needs to a cloud provider such as Amazon Web Services and Microsoft Azure. Cloud computing enables its customers to use computing resources as a service, and pay only for what is used. By leveraging cloud computing, organizations hope to optimize costs, and increase their computational capabilities, without having to purchase and manage the necessary hardware and software.

Cloud computing also democratizes access to computing resources. For example, software developers in small organizations are empowered to launch globally available apps and online services – something that would have been highly unaffordable if they had to purchase and own massive computing resources. Similarly, researchers with low or moderate resources can share and analyze data at scales once possible only for researchers with access to deep pockets. Finally, end-users in society can quickly access software and storage to make, share, and store digital media in quantities that reach far beyond the computing capacity of their personal devices.

Cloud computing is essentially the technology that allows computing resources to be utilized as a service, as opposed to traditional "ownership model" of those resources. The National Institute of Standards and Technology defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [Bohn 2020].

Cloud computing vendors like Amazon Web Services and Microsoft Azure allow users to select the service configurations that they currently need, and only pay for those; hence providing a customized and cost-effective solution from a large menu of available configurations. The Cloud vendor then "provisions" the customer's requested configuration. The Cloud customer also has the ability to rapidly change their configurations on-demand, and add on or release computing resources depending on their own current needs. In addition, the Cloud customer can also customize the configuration to their own specific needs, especially if the provisioned configuration has to work with their own existing legacy systems, while also requiring minimal regular management. We next discuss the various cloud computing delivery models.

The primary purpose of this paper is to provide a comprehensive understanding of the cloud computing environment and is organized as follows. We start by describing the various Cloud delivery and deployment models. The paper then investigates and identifies the major advantages and disadvantages of cloud computing. We then present some cloud computing case studies and present the major cloud service vendors and their cost structures. The paper then discusses some future trends in cloud computing and ends with a summary.

CLOUD COMPUTING DELIVERY AND DEPLOYMENT MODELS

Cloud Computing Delivery Models

Cloud Computing is often delivered using three delivery models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [Glass 2020]. Some vendors like Microsoft Azure and AWS are able to deliver all types of delivery models, while some only deliver a single type. The SaaS delivery model allows end-customers to skip the process of purchasing, installing and maintaining software on premises, and allows the use of software on a "pay-for-use" basis. For example, customers can subscribe to Microsoft Office 365, and pay a monthly charge for using the Office software as a service. Other examples are Netflix, in which, video streaming is the subscribed service or Dropbox, where storage is the subscribed service. SaaS is often implemented as a multi-tenant system. where all the users use a single version of the software. Many SaaS based products are user friendly and can be used from web browsers or SmartPhone Apps.

PaaS deployment allows the customer to host its own applications on a Cloud vendor's platform without getting involved with the background hardware and software resources needed to run the application. Unlike SaaS, PaaS users need software developers to develop and maintain their own application and data, but not worry about provisioning to the end users. The PaaS vendor provides a robust hardware and software platform for developing, controlling versions, testing, running, and hosting applications. The in-house software developers can just focus on the application development aspects, using the PaaS provided services, to meet the needs of their end users – thus making it cheaper and faster for the developers.

The PaaS market is growing rapidly at a CAGR of about 13%, [Kanade 2022] and is expected to be about $176 billion in 2024 [Vailshery 2024]. Examples of PaaS providers are Microsoft Azure App Services, Amazon Web Services (AWS) Elastic Beanstalk, Google App Engine, SAP Cloud and Salesforce Lightning. PaaS platforms are well suited for the modern organizational environment, where application development and deployment need to be rapidly adapt to changing conditions and user expectations. PaaS customers are usually enterprises, but can be individuals too. An interesting example of a PaaS use case is GitHub Copilot (GC), which is a cloud-based AI-based code writing service, and they market it as "your AI pair programmer." GC is based on OpenAI's Codex, and is able to suggest code and entire functions in real time. The GC PaaS is hosted on Microsoft Azure, and is priced at just $100/year for individuals, and $19.99 per user per month for businesses. GC for individuals was launched only in June 2022, and at that time 27% of developers' code files on average were generated by GitHub Copilot, while in Feb 2023, that number rises to 46% across all programming languages. For Java, that number in 2023 was a whopping 61% [Zhao 2023]. Developer productivity has clearly shot up rapidly with the GC PaaS.

An IaaS delivery cloud platform, as its name suggests, is the on-demand provisioning of computing infrastructure for its customers, which are also typically enterprises, including research organizations. IaaS allows the users to have a lot more control (and hence also needs more in-house technological work), and essentially expects the IaaS provider server farms to create Virtual Machines (VMs) for the users. The VMs are housed by the IaaS vendor, and all aspects of servers, hard drives, networks, and load balancing are handled by them. The user still has the responsibility of choosing the Operating System and developing application software, but is able to quickly provision, deploy and scale the applications due to the IaaS vendor services. Amazon Web Services, Rackspace, Digital Ocean, Google Cloud Platform, and Alibaba Cloud are examples of IaaS vendors. While IaaS gives the most flexibility, it also requires more user involvement and responsibilities. Research facilities needing high computing using supercomputers or clusters will benefit by using a scalable IaaS vendor. Big Data analysis also is a great use case for IaaS as it needs large workloads and a lot of processing. GE Healthcare, for example, uses AWS to collect, store, process and access worldwide – thus allowing its customers to obtain value from nearly a petabyte of health imaging data [Hoffman 2023]. As seen above, depending on the needs of the customer, Cloud vendors offer various types of delivery models. Another dimension of clouds is the type of deployment, as discussed next.

Cloud Deployment Models

A Cloud deployment model classifies the type of cloud environment depending on factors such as ownership, location, access, and management responsibilities [Robinson 2024]. As Figure 1 [Peterson 2023] shows below, the four types of Cloud deployment models are Private, Public, Hybrid, and Community. We next briefly discuss these cloud deployment models.

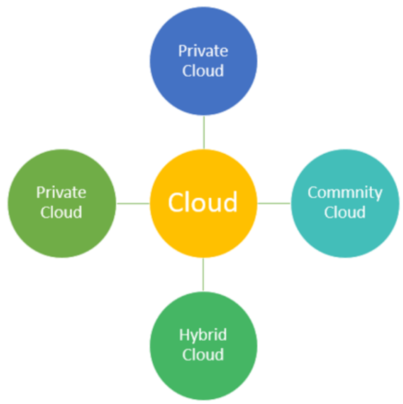

Figure 1: Cloud Deployment Models

Private Clouds are owned and operated by a single enterprise, and provides the information service needs for its departments, users, customers, and partners [Tavbulatova 2020]. Private clouds allow for a lot of customization and control, but also requires the enterprise to invest heavily in the entire Cloud deployment, including its security. Figure 2 [Peterson 2023] shows the general Private Cloud architecture. Private Clouds are used more often by relatively larger enterprises as it does need high capital and operational expenditures.

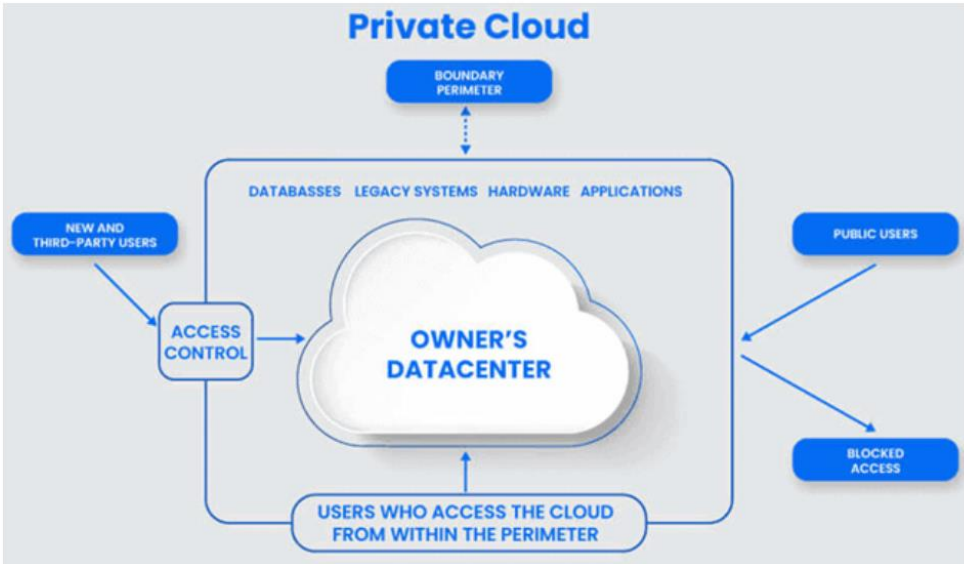

Figure 2: Private Cloud Architecture

Public Clouds are designed to provide computing resources to the general public or a group of companies, usually via the worldwide web, and uses a public Cloud vendor, such as Amazon Web Services or Microsoft Azure. The vendor is responsible for managing and maintaining the cloud, thus reducing the burden on the users. The cloud users typically use a self-service portal to use and manage the resources provided by the vendor [Tavbulatova 2020]. Public clouds are relatively simple to use, and are cost-effective for even small or medium sized businesses. They can provide unlimited quantity of computing resources (scalable), provide high data security, and allow for rapid system implementation and updates. However, it requires reliable, high speed internet connectivity, and limits the control of cloud operations. Complete dependency on the vendor system (including security vulnerabilities), and the higher risk of vendor lock-in are potential shortcomings [Golightly et al. 2022]. Figure 3 [Peterson 2023] shows the general Public Cloud architecture.

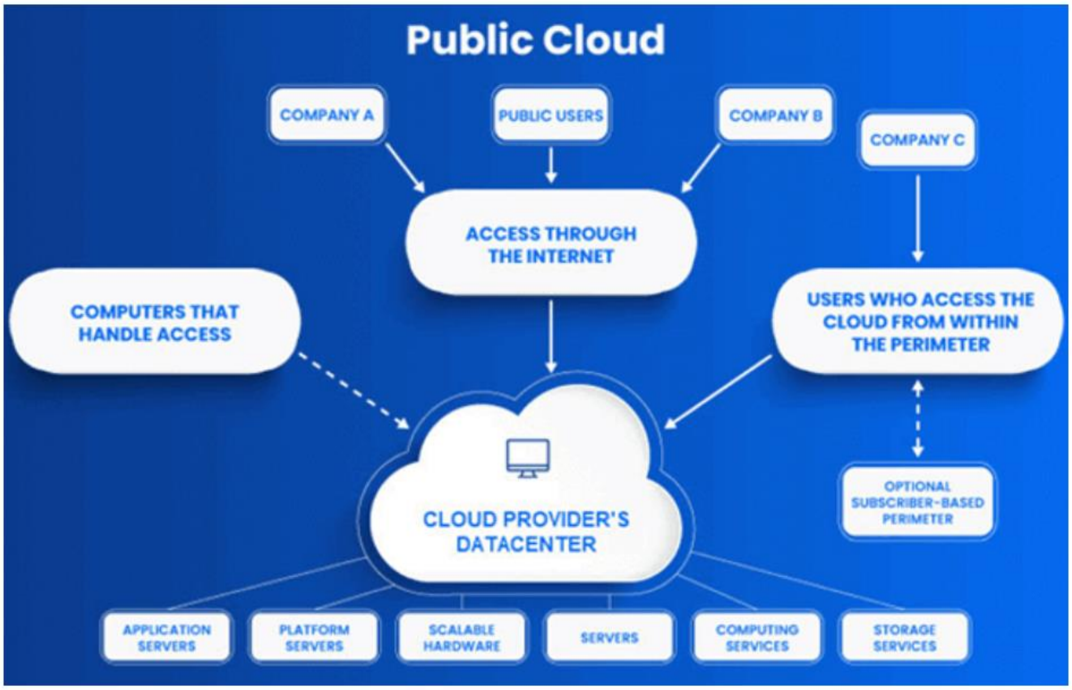

Figure 3: Public Cloud Architecture

Hybrid Clouds are a combination of public and private cloud models. The enterprise may need complete control over a portion of the system with sensitive data or mission critical applications, which can then be deployed on a Private Cloud and maintained in legacy systems. A Public Cloud Vendor can then be used for all other uses. The two types of clouds, along with any legacy systems, will then need to be integrated. Figure 4 [Peterson 2023] shows the general Hybrid Cloud architecture. Hybrid Clouds enable users to capitalize the flexibility of the Cloud, while still realizing value from traditional infrastructure, can maintain data security, reduce costs by transferring resources to cloud providers, increases the available processing power with provider resources, and moves loads from the local infrastructure to the cloud and back [Tavbulatova 2020, Golightly et al. 2022, Zhukovskiy 2017]. Management of Hybrid Clouds is also more complex [Peterson 2023].

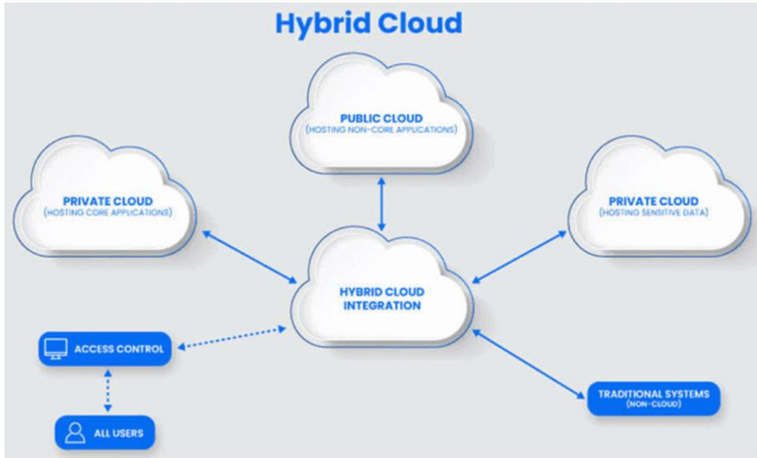

Figure 4: Hybrid Cloud Architecture

Community Clouds allow a group of organizations to share the resources and services using one of the aforesaid architectures (private, public or hybrid), but based on mutually agreed upon regulatory requirements. This Cloud computing model is operated and managed by third party vendors, community members or both [Peterson 2023]. The members of the community usually share similar security, privacy, performance, and compliance requirements. The community clouds allow remote storing and accessing of data/files on different computers and makes the data available anywhere in the world with low cost for data utilization [Tavbulatova 2020, Golightly et al. 2022]. A major drawback is security and segmentation because many different communities are involved [Peterson 2023].

Finally, it is worth mentioning that some enterprises choose multi-cloud deployment, in which it chooses to use multiple public Cloud vendors. This might be desirable to prevent getting locked-in to a particular vendor, better negotiation for pricing and features, and picking the best vendor for each feature. This approach does add to the complexity of Cloud services management, and it also requires personnel who are competent with products from various Cloud vendors. [Peterson 2023]. Table 1 [Peterson 2023] compares the strengths and weaknesses of the various Cloud deployment models.

Parameters	Public	Private	Community	Hybrid
Setup and use	Easy	Need help from a professional IT team.	Require a professional IT team.	Require a professional IT team.
Scalability and Elasticity	Very High	Low	Moderate	High
Data Control	Little to none	Very High	Relatively High	High
Security and privacy	Very low	Very high	High	Very high
Reliability	Low	High	Higher	High
Demand for in-house software	No	Very high in-house software requirement	No	In-house software is not a must

TABLE 1: Comparing Cloud Deployment Models

ADVANTAGES OF CLOUD COMPUTING

There are many efficiencies and advantages to be obtained through the implementation of cloud computing, but the technology does not come free of potential issues or disadvantages. Every organization or consumer must critically analyze their specific needs, desired outcomes, and sensitive areas in order to determine whether or not the implementation of cloud computing is a strategically sound decision.

When looking at the possible advantages of cloud computing, many firms' initial interest is to gain operational efficiencies and reduce costs. Cloud computing has the ability to not only reduce operating costs, but also increase flexibility, reduce down time, quicker time-to-market, decreased deployment costs for new IT services, and increase security [Webroot 2022; Danave 2024]. Cloud computing allows for much seamless updates to software and hardware, as the cloud operators can do this more efficiently, and at a lower cost. Cloud computing can also eliminate costs in real estate, operations of facilities, personnel, and utilities required to house and maintain/manage servers and data storage centers. Crucially, less IT staff need to be hired and maintained on the payroll, while still having easy access to the latest technical skills from the cloud provider.

Many organizations will find that the costs associated with purchasing some form of cloud computing can be much lower than the costs they previously incurred when operating these technologies in-house. This also creates an advantage in terms of scalability. Not all firms have the capital or real-estate required to scale up their operations in terms of computing power or data storage. The Cloud offers virtually endless space to scale up operations at a more affordable rate. This is especially critical to smaller firms seeking to scale up their operations quickly and efficiently. Cloud computing also offers an advantage in terms of mobility. Essentially, cloud computing allows organizations to be more efficient, reduce costs, stay more flexible – all while accessing their Cloud based resources anywhere in the world.

A big advantage of Cloud computing is the ability to control costs, capacity, and performance on the fly. The auto-scaling feature of many Cloud services (like Amazon's AWS) allows user applications to handle varying loads in real-time. This allows Cloud customers to only pay for resources being used, and handle varying loads. For example, when the popular augmented-reality game "Pokeman Go" was launched on Google Cloud, the game designers expected a worst-case scenario of 5x expected traffic load. The actual load turned out to be much greater at 50x the expected traffic load [Shivang 2022]; but the system worked flawlessly due to the auto-scaling feature. Google Cloud auto-scaler uses the Google Kubernetes Engine and Cloud Spanners to achieve this [Vergadia 2021].

It is worthwhile to see the evolution and advantages of Kubernetes with respect to cloud deployment. Initially, each server ran one application, so it was secure but very inefficient due to a lot of idle time, needing many servers (and needing more personnel, more real estate, more electricity etc). Virtualization, and the creation of Virtual Machines (VMs) was the next stage of evolution. In this each physical machine/server could have many independent applications and also operating systems. This considerably reduced the number of servers needed and increased utilization rates and reduced costs. But as Cloud computing came into the IT scene, it would be nice to decouple the application in each virtual machine from its own operating system, and low-level infrastructure so that those functions could be shared. This new decoupled virtual machine system was called containers. Containers are similar to a VM as it has its own filesystem, share of CPU, memory, and process space, but they are decoupled from the underlying operating system, so they are portable across clouds and OS distributions. Containers can be easily deployed, moved, deallocated, and reallocated in different computers and operating systems – in a dynamic way. These container advantages are very suitable for running on clouds or be portable across computers with different operating systems. In systems that need little downtime, it would be nice to quickly get another container working, if a container failed. This is where Kubernetes comes in – it is an open-source platform developed by Google to manage containers. Kubernetes can automatically run distributed systems with resiliency, allowing containers to be deleted and created over a cloud, making load balancing and scaling even more efficient.

The advantages and nature of cloud computing described above allow companies to use them for business continuity and disaster recovery. A local disaster will be a lot less disruptive if the business location faces a disaster, but had all or most of its computing infrastructure in a cloud. Another possible application would be collecting data from Internet of Things (IoT) sensors and applications. The sensors can be distributed over a large geographical area and collected a stored in the cloud. As the data volume increases or decreases, Cloud resources can be dynamically allocated.

Finally, a major benefit of migration to cloud based infrastructure is environmental sustainability [Webroot 2022]. The operational efficiencies mentioned above are partly due to the energy savings of using cloud-based servers that are typically newer and more energy efficient. Another reason is that many cloud data centers employ carbon offsets or use renewable energy source for their data centers, so using the cloud vendors automatically offsets some of the carbon used by IT. For example, Google Cloud plans to be 100% carbon free by 2030 [Lardinois, 2021], and Google offers a free feature that provides its users with custom carbon footprint reports that detail the carbon emissions their cloud usage generates.

DISADVANTAGES OF CLOUD COMPUTING

As seen above, Cloud computing offers compelling potential advantages to many organizations, however, for some organizations, the potential disadvantages are more crucial. One of the concerns involves downtime and loss of control. When a company decides to use Cloud computing services, the infrastructure of the Cloud is owned, managed, and monitored entirely by the Cloud provider. Downtime and lack of control over that is a major concern. Since Cloud computing systems are all internet based, service outages are always a possibility that could occur at any time. In 2017, Amazon Web Services had an outage that cost publicly traded companies up to 150 million dollars [Larkin, 2020]. Outages and slowdowns are not something that companies can control, but the company could incur a major cost due to this. For companies to mitigate the risk that comes with outages, they have been designing their services with disaster recovery in mind. Services such as AWS Direct Connect, Azure ExpressRoute and Google Cloud's Dedicated Interconnect or Partner Connect are options that companies can implement. These services provide a network connection that is dedicated to keeping them online and reducing the risk of business interruption from public internet outage. Some organizations may just not want to put themselves in this vulnerable situation, and avoid Cloud services.

Another negative aspect is the potentially high costs of Cloud Computing – for some organizations, continuously paying for access to Cloud computing can eventually add up to a cost greater than if they would have purchased and installed the technology themselves on premise. In terms of costs, cloud computing can be a balancing act depending on the size and nature of the utilized technology.

Another major concern for Cloud users is cybersecurity, as hackers can potentially compromise the security of the Cloud vendor's data, software, and hardware. This is especially true in multitenancy, where many organizations share common resources of the Cloud provider. There is also the possibility for Cloud service providers to misuse the customer's data through means that were not previously agreed on. It is interesting that many online users have just given up on privacy and security. As a result, many are numb to the issue of data protection on the cloud or otherwise. About 20% of users don't care how much data they share online, while 26% believe it's "inevitable" that their data will be leaked so they don't worry about it [Gregalis 2022]. A recent study [Gregalis 2022] revealed that many adults shared private secrets on a cloud messaging service. About 16% shared their sexual fantasies, 14% shared sexist, racist, homophobic, or otherwise offensive comments. In addition, about 12% of the respondents admit to sharing details about substance abuse, while about 10% reported that they cheated on their partners.

What is surprising is that many share so much of their private data despite knowing the many damaging consequences of unsecured data. That same research [Gregalis 2022] also reported that 79% of those who have discussed private topics admit they could face serious consequences if their discussions online were leaked. Almost 50% admitted that any compromised data would ruin their relationships with friends or family, 28% would be left open to blackmail, 22% could lose their job, 19% their partners, or 10% even custody of a child. As expected, the younger generation of online users is less concerned with data being leaked; especially about half [Gregalis 2022] were unconcerned with financial theft due to leaked data, though about 15% claimed that they were concerned with their online reputation.

A recent study [Ponemon 2022] surveyed 1,500 IT and security leaders, and reported that 60% of them are not confident in their organization’s ability to ensure secure Cloud access, even as adoption continues to grow across a diverse range of cloud environments. Organizations face several barriers to securing their cloud environments, with the top challenges being network monitoring/visibility at 48%, in-house expertise at 45%, increased attack vectors at 38% and siloed security solutions at 36%. Many (62%) also indicated that traditional perimeter-based security solutions are no longer adequate to mitigate the risk of threats like ransomware, distributed denial of service (DDoS) attacks, insider threats and man-in-the-middle attacks.

The healthcare industry is a big user of Cloud services as it allows users to access data from multiple devices from anywhere that has internet, making traditional methods of storing data a thing of the past. The downside of cloud computing is the increased risk of being hacked or stolen by hackers. So, it is very crucial to only use HIPAA compliant cloud computing services. In healthcare, safeguarding protected healthcare information and abiding by HIPAA compliant cloud data storage requirements is a mandatory requirement. As a result, Cloud vendors are constantly looking for ways to make cloud-based system more secure. Things like improving the screening and hiring practices of employees and privileged users is also important, as not everyone should have this unlimited access.

While security technologies continue to advance, so too do the technologies used by hackers with malicious intent. Cloud service providers implement some of the best security standards, but there is always a risk when storing data on external services. For companies to deal with this limitation, they must understand the shared responsibility model of their cloud provider, and be aware of what their cloud provider covers in case of a security breach. For a company to reduce the risk of security, it is important for them to implement a risk-based approach with security and make security a core aspect of all IT operations. Cybersecurity threats will always exist, and this threat and lack of control on Cloud platforms, is often enough to deter some organizations from utilizing public Cloud computing.

Another big disadvantage for global organizations and nations is related to security of data and intellectual property housed in a Cloud Platform in a foreign country. Geopolitics and National security might also prevent certain organizations from using foreign Cloud vendors, or might require a global organization to store the country data within the country. An interesting example is Tiktok, the popular social media for sharing videos, which belongs to a Chinese company called ByteDance. Tiktok initially stored all US user data in private servers in Singapore and Virginia, but the US Government is worried that the Chinese government might summon the US data from ByteDance and pose a national threat. The data can also be potentially used for targeting misinformation to US citizens that harms USA and its citizens. USA then threatened to ban TikTok, unless all its data is stored in Infrastructure of a USA company and prevent transporting that data outside the country. Tiktok was then forced to spend billions of dollars by moving all data to Oracle Cloud servers and instituting protection so that the data cannot be transported to other countries [Fung 2022].

Vendor lock-in is another limitation of cloud computing. Once a company chooses to move to the cloud, it gets increasingly harder to get out of cloud-dependence [Shapira, 2021]. Switching between cloud services of different vendors has not become an easy or inexpensive process yet. The process for a company to switch from one cloud vendor to another is often costly and burdensome. The steep "switching costs" and vendor-stickiness could, in turn, incentivize cloud vendors to increase fees for customers over time, because the costs and time needed for switching to another cloud vendor are great barriers. There are many costs to consider such as Egress Fees, the cost of adding new services in the future and worse, lock one out of new features available with other vendors. [Banthia 2024].

Some services, like Amazon's AWS long-term Glacier service, even monetize their lock-in. While the company does charge a monthly storage fee based on how much users are storing, they don't charge anything for uploading data to the service. Users can upload a gigabyte or 100 terabytes and still pay the same $0.00 transfer fee. But data transfer above 1GB/month out of Glacier to the rest of the internet is charged, ranging from $90 per terabyte transferred down to $50 per terabyte, depending on how much users are moving. Even though it is not a tremendous fee for the amount of data, it still shows how Amazon wants to reduce the friction of uploading data and increase the friction of moving it back out. [Gewirtz 2021].

Due to the difficulties of switching vendors, there are now some Cloud Migration Services vendors that help users migrate services across cloud platforms. Capgemini, Accenture, Deloitte, and InfoSys are some of the leading cloud migration service vendors [Hein 2022]. The Cloud Migration Industry is poised to grow by USD 7.09 billion between 2019 and 2024, resulting in a 24% annual growth rate during the forecast period [Technavio 2022]. Another risk associated with changing cloud vendors is the possibility of exposing sensitive or proprietary data during the migration process. For a company to mitigate the risk of vendor lock-in and cloud migration, companies should design their cloud architecture with best practices. Cloud migration service vendors may also be very beneficial for risk mitigation.

It is up to each firm or consumer to determine whether the potential advantages outweigh the potential disadvantages when making the decision to implement cloud computing – whether it be outsourced or housed on-site. As stated above, Cloud computing has been shown to save time and money, as it is an efficient process. Although it has many benefits, cloud computing also has its limitations: Downtime, Security and Privacy, Limited Control and Flexibility, Vendor Lock-in, and Cost. Despite the preeminence of cloud architecture, we've seen numerous companies drop public cloud, whether partially or in full, in favor of setting up their own infrastructure with reasonable success. Dropbox is the highest-profile example, saving some $75 million over two years after bidding farewell to the Amazon cloud. And it's not alone in this push [Shapira 2021].

CASE STUDIES

Lenovo's use of Microsoft Azure illustrates a great use of Cloud Services, even for large companies. Lenovo builds millions of computers for customers every year, and wants its customers to trust that each one arrives configured to the customer's request. To provide this attestation, the company tracks each new device with its Trusted Supply Chain system, built with ledger in Microsoft Azure SQL Database. Lenovo chose the ledger in Azure because of the performance, scalability, and security of the Azure cloud platform. Lenovo is so pleased with the results that it is planning to use ledger in Azure SQL Database for additional high-security applications [Azure 2022]. Thorsten Stremlau, Lenovo's Chief Technologist and Executive Director of Commercial Product Portfolio, says that "Microsoft has extensive certifications in place for its own infrastructure, and we've seen more and more of our corporate customers moving to Azure as their trusted cloud platform," and adds that "the scalability of the Azure pricing model is also attractive and better than the competition."

Another interesting case is how Twitter migrated from an on-premise architecture to Google Cloud Platform to boost the reliability and accuracy of Twitter's ad analytics [Phalip 2020]. Twitter, as part of the daily business operations on its advertising platform, serves billions of ad engagement events, each of which potentially affects several downstream aggregate metrics. To enable its advertisers to measure user engagement and track ad campaign efficiency, Twitter offers advertisers a variety of analytics tools, APIs, and dashboards that can aggregate millions of metrics per second in near-real time. Twitter originally operated its Ad Analytics applications and data on its on-premise data center architecture. With time, Twitter's system was reaching its performance limits, getting harder to add or change functionality, and getting expensive to operate. The system was also prone to sporadic failures. So, Twitter decided to rethink the architecture and deploy a more flexible and scalable system in Google Cloud. Twitter began seeing huge benefits in just six months, after transitioning its ad analytics data platform to Google Cloud. Twitter's developers gained much agility, as they could easily configure existing data pipelines and build new features much faster. The real-time data pipeline has also greatly improved its reliability and accuracy, thanks to features of the Google Cloud Platform.

A recent interesting use case of Cloud Computing is the deployment of OpenAI's APIs for ChatGPT and Whisper. The ChatGPT API runs as a service via Microsoft's Azure on a Public, multitenant Cloud. Many users from different organizations will be sharing the same hardware, and hence the cost would be 10 times lower than the earlier version [Deutscher 2023]. This would likely allow many software developers to now incorporate ChatGPT based functionality in their software. OpenAI is also deploying a Whisper Transcription Service API via the Cloud, which would allow developers to deploy software with automatic transcription in either the original language, or translate it to English.

A final case illustrates how the PBS (Public Broadcasting Service) uses Amazon Web Services (AWS) Cloud Platform to efficiently stream videos [AWS 2020]. PBS has over 330 member stations and reaches over 100 million people through television, and 32 million people online monthly with diverse education and entertainment programming and applications. PBS uses Amazon Web Services (AWS) to bring its annual Short Film Festival to an online format, thus improving the performance of video streams. Since using Amazon CloudFront content delivery network, PBS has experienced 50 percent fewer streaming errors. PBS delivers nearly all of its video streaming through CloudFront, averaging more than 70 PB of content delivered monthly. In addition, PBS enhanced its archives with deep search functionality using machine learning and artificial intelligence (AI) – also running on the cloud. In summary, Amazon cloud services innovated PBS's backend operations and content delivery and functionality and drove forward its mission of being a trusted window to the world.

PRICING AND VENDORS

Pricing for cloud computing can vary depending on how much storage a company plans on using. The 'big three' commodity clouds are: Amazon AWS, Google GCP, and Microsoft Azure. Below is a summarized breakdown of their pricing structure.

Amazon AWS offers three different categories for their cloud pricing benefits: Pay-as-you-Go, Reserved Capacity, and Volume-Based Discounts. Pay-as-you-Go option allows customers to only pay for the services that they use and when they use the service. The Reserved Capacity is a very common cloud concept. This concept allows customers to reserve the amount of compute or storage that their company will need ahead of time, and this will save the company as much as 75% over the on-demand pricing. The last option is Volume-Based discounts based on economies of scale. The pricing is as follows. Up to 50TB of storage is 0.023 GB/month, while 500TB+ will reduce the cost to 0.021 GB/month. The more storage customers use, the less they will pay for each GB used [Gil 2022, Faddom 2022].

Google GCP pricing is also built around the pay-as-you-go model but promises no activation or termination fees. Google GCP offers four different pricing structures. GCP offers Preemptible VM Instances, this is for companies whose workload doesn't need a steady availability. This allows the company to set up their database to be interrupted when necessary, and saves the company up to 79%. The second pricing structure is Per-Second Billing, this allows companies to pay for exactly what they use by the second. Third is, Sustained- Use Discounts. This pricing structure decreases the rate with increased workloads running. Lastly, offered by GCP is Committed-Use Discounts [Faddom 2022].

The last of the 'big three' in the cloud computing world is Microsoft Azure. Azure's cloud pricing comes with the ability to reserve GB ahead of time, but in order to get price savings customers must have a one-to-three-year commitment. Azure allows a company to pay less for development and testing resources by including no software charges. Azure is most cost efficient when using Reserved capacity versus pay-as-you-go on Azure.

It is very important to manage costs in a cloud computing environment, which has spawned a whole new area called "Cloud Financial Operations," often referred to as just FinOps [Robinson 2024]. FinOps is the practice of bringing a financial accountability cultural change to the cloud computing, enabling distributed engineering, finance and business teams to make trade-offs among speed, cost, and quality in their cloud architecture and investment decisions. The FinOps framework developed by the FinOps Foundation [Finops 2024] is a structured way to tame the rising costs of cloud computing and maximizing its value to an organization.

FUTURE DEVELOPMENT AND TRENDS IN CLOUD COMPUTING

The future of cloud computing definitely looks very bright, as it is expected to significantly change the way businesses operate - with over 90 percent of businesses utilizing it for multiple reasons including, data backup, email, and disaster recovery purposes just to name a few [Wiggins 2022]. Cloud computing will give businesses the ability to create more personalized services for their customers and provide those services at a greater speed and lower cost. By tapping into the cloud, businesses will become more scalable as they would be able to service more customers. The cloud market is projected to grow from above $500 billion in 2022 to about $1.25 trillion in 2027, while storing 200 zettabytes (2 billion terabytes) in the cloud by 2025 [Griffiths 2023].

As the uptake of Cloud Computing marches on, there are several promising trends to take note of – ones which will make Cloud Computing even a more compelling technology for organizations. The major ones discussed next include Multicloud, Edge Computing, Artificial Intelligence, Serverless applications and better Cloud security.

Multicloud is when an organization utilizes two or more Cloud providers to run their applications [Danave 2024]. Cloud vendors can differ in pricing, capabilities, customer support, etc., so the ability to choose multiple vendors allows an organization to tailor each app to match the best Cloud provider. It also allows organizations to combat the major problem of vendor lock-in. Multicloud deployments also work well with DevOps development practices and cloud-native portability enhancing microservices architectures such as containers. Microservices allow applications to be split up into smaller independent parts such as containers. A single user query to a microservice based application can trigger a group of internal microservices (say, containers) and then stitch them together to compose a response to that user query. Furthermore, Multicloud deployments built on top of technologies such as Kubernetes (which was discussed before) allows for greater flexibility and portability across multiple Clouds and computing environments. So, it is getting easier to mix and match the services of different Cloud vendors.

Edge Computing, along with 5G cellular, IoT and Cloud Computing networks is another development that adds considerable value to IT eco systems. The proliferation of IoT devices such as sensors, wi-fi enabled devices like lawn irrigation, home automation, farm automation, factory automation etc. generates a lot of data that can be processed near the source of data collection using edge devices. The processed data (reduced, summarized data) then can be sent to data centers on the Cloud. Another example of using edge computing with cloud is streaming videos – a streaming service such as Netflix has high demand for newly released, popular content. So instead of streaming that popular content to every user from the central server housed on a Pubic Cloud such as Amazon's AWS, many temporary copies can be kept in several small edge devices (also known as micro data centers) in several locations in many cities, and streamed to the homes close to that micro data center. Advantages will be lower latency (lower delay) of high quality content, less network congestion on the overall network connecting all network users. Edge computing's main advantages are to process user requests as close to the data as possible, and reduce latency. If there is no Internet connectivity, these edge devices can transmit the data using 5G cellular networks that also have very low latency. Overall edge computing, along with Cloud Computing adds to operational efficiency and performance enhancement. This will continue to help with reducing network bandwidth requirements, and is expected that 2024 will be the year in which there will be seamless integration of edge computing with cloud services [Danave 2024].

Artificial Intelligence is another field that is increasingly being deployed to enhance Cloud computing. The first area of Cloud improvement is in the automation of Cloud efficiency and management in routine activities such as streamlining operations, provisioning, load balancing, self-healing, and security. The increased efficiencies clearly result in lower costs for the Cloud provider, and also makes the Cloud more responsive to customer needs, while keeping costs down. Cloud vendors can also offer AI as a Service (AIaaS) to customers. Organizations of all sizes can have easy and cost-effective access to AI to enhance their own operations. Cloud customers may use the AI for data analytics, predictions, or better customized experience for their own customers without investing large amounts for building their own AI systems. A Deloitte study found that 70% of companies got their AI capabilities through cloud-based software, and 65% create AI applications using cloud services [Brenner 2023]. Another example is the inclusion of AI-based “Einstein” in the CRM Cloud software provided by Salesforce. Einstein uses AI to analyze customer sales data and provide actionable tips that potentially enhance the customers’ current sales strategies [Brenner 2023]. In summary, AIaaS offers organizations pre-built AI models, tools, and APIs hosted on cloud computing platforms. This enables organizations to seamlessly implement AI functionalities, even without specialized AI expertise and infrastructure [Danave 2024].

Serverless applications are another exciting development in the evolution of Cloud computing. Serverless applications are hosted on a Cloud and allow application developers to write code, but do not worry about the backend Cloud infrastructure required to run the applications. All the backend functions like allocating servers, memory, security etc. are completely handled by the Cloud provider. A very popular serverless application architecture is Function as a Service (FaaS). The developer writes the application as a collection of functions, each of which is triggered by some action, such as clicking on a weblink. The functions are loaded to the Cloud provider, and the functions Run on the Cloud, when they are invoked. The Cloud provider allocates all resources needed to execute the Function. In traditional cloud computing, the user still has to estimate the size of servers needed, space needed etc. and pay a monthly subscription for the chosen options, with an option to scale up dynamically. Typically, users pay for more than they need so as not to pay up for extra loads, but risk paying for a lot more resources than they really need, at normal times. In Serverless model, servers are still needed, but all the allocation, scaling etc. is completely done dynamically by the Cloud provider, and the user only pays for the resources used to run the functions, after which there is no other cost. This setup is perfect for functions that are used for short durations by few users, occasionally. The user only pays for resources, when they are used, after which the servers can be used by other Cloud users, resulting in lower user costs.

Serverless applications scale automatically, and do not even need other management platforms like Kubernetes for Container management (which was discussed before). Developers of these serverless applications can mostly focus on the code development and not worry about any hardware issues, so their productivity and update cycles are greater. There are disadvantages to this serverless approach such as latency, security and debugging [Hadidi 2022]. When serverless functions are triggered, it takes some time for the Cloud provider to set up all the needed resources. This delay is noticeable when the function is called for the first time or after a certain long time period during which the function is not used. The Cloud provider deallocates all resources if the function is not triggered for a while, so that the hardware can be used by other users. That also leads to possibly less security, as the same server is used by several users and, if not properly configured, data could be leaked to other users. Another current problem is the lack of visibility of the backend after deployment, and the added complications to debugging the code on the Cloud Platform. In summary, serverless applications, in its current form, are particularly suited when applications can be designed as a collection of relatively decoupled functions, that do not require a relatively large compute workload over an extended period of time. Serverless Computing is growing at a blazing Compound Annual Growth Rate (CAGR) of 23.17% between 2023 and 2028 [Danave 2024].

Finally, as described in the disadvantages section above, security is always a major concern for organizations deploying their mission critical systems on a Cloud. The evolving Zero Trust Network Access (ZTNA) [Ponemon 2022] can address many of these security challenges, and also accelerate the adoption of cloud services. ZTNA is an IT security solution that is more secure than traditional VPNs. VPNs provided secure access to entire networks, but was unable to control access within the network. ZTNA provides superior secure remote access to an organization’s applications, data, and services using well defined access control policies. Each user or role can have access only to specified IT resources (such as cloud services or applications) in the network. As organizations adopt multi clouds, ZTNA provides secure, robust, and well-controlled access that doesn’t assume any implicit trust but creates an identity-based, context-based, logical–access boundary to applications. ZTNA can be implemented as an in-house stand-alone product for cloud-averse organizations or as a service on Clouds. According to Gartner, almost 90% of organizations use ZTNA-as-a-Service on a Cloud platform. Some of the top ZTNA vendors are Palo Alto Networks, Z-Scalar, Cloudfare, Akamai and Cisco.

As cloud-native development practices grow, 90% of respondents will have adopted DevOps, and 87% will have adopted containers within the next few years, but modern Cloud security practices aren’t as widespread. For instance, only 42% can confidently segment their environments to apply the principle of least privilege, and nearly 33% of organizations have no collaboration between IT security and DevOps, thus presenting a significant risk. Respondents that have adopted a ZTNA strategy report 65% more productivity of the IT security team, 61% stronger authentication using identity and risk posture, 58% increased productivity for DevOps and about 58% increase in network visibility and automation capabilities, as the top benefits [Ponemon 2022]. In the next year or two, significant advances in ZTNA is expected as established cloud vendors acquire promising ZTNA startups and integrate ZTNA features in their existing platforms [Ray 2024].

SUMMARY & CONCLUSION

Cloud computing is rapidly reshaping the computing landscape, offering organizations and individuals alike unprecedented access to resources and agility. By shifting from on-premise infrastructure to cloud-based services provided by giants like AWS and Azure, businesses gain strategic advantages. They optimize costs, scale computing power on demand, and free up internal resources to focus on core competencies and customer needs. Cloud computing not only streamlines operations but also democratizes access to computing resources, enabling small organizations, researchers, and end-users to benefit from global-scale applications, data analysis, and digital media storage.

In conclusion, this paper provides a comprehensive understanding of Cloud computing, spanning its nature, architecture, and diverse delivery models. Through insightful case studies, it highlights the real-world applications of cloud technology, illuminating both advantages and disadvantages. A critical analysis of major cloud service vendors and associated costs deepens the understanding of the landscape. Furthermore, the exploration of future trends in cloud computing sheds light on potential developments. The future of computing undoubtedly lies in the Cloud. As technology continues to evolve, embracing the Cloud's potential will be crucial for staying competitive, fostering innovation, and unlocking limitless possibilities.

REFERENCES

Anand, Binod; "Cloud Computing Future: 12 Trends & Predictions About Cloud," https://www.knowledgehut.com/blog/cloud-computing/cloud-computing-future; July 2023

AWS, "PBS on AWS," https://aws.amazon.com/solutions/case-studies/innovators/pbs/; 2020

Azure, "Lenovo reinforces customer trust with ledger in Azure SQL Database," https://customers.microsoft.com/en-us/story/1497685499820529889-lenovo-manufacturing-azure-SQL-database-ledger; May 2022

Banthia, Bakul; "Understanding the Risks of Cloud Vendor Lock-In," https://drj.com/industry_news/understanding-the-risks-of-cloud-vendor-lock-in/; August 2024

Bohn, Robert, Lee, Craig; Michel, Martial; *NIST Cloud Computing Program – NCC;* 2020 February 13, https://www.nist.gov/programs-projects/nist-cloud-computing-program-nccp

Brenner, Michael' "The Role of AI in Cloud Computing," https://www.nutanix.com/theforecastbynutanix/technology/ai-in-the-cloud; July 2023

Danave, Sameer; "Top 5 cloud computing trends of 2024," https://www.cncf.io/blog/2024/05/03/top-5-cloud-computing-trends-of-2024/; May 2024

Deutscher, Maria; "OpenAI makes its ChatGPT and Whisper models available via cloud APIs," Mar 2023; https://siliconangle.com/2023/03/01/openai-makes-chatgpt-whisper-models-available-via-cloud-apis/

Faddom, "Cloud Computing Costs & Pricing Comparisons for 2022." https://faddom.com/cloud-computing-costs-and-pricing-comparison/; 2022

Finops Foundation, "What is FinOps?," Retrieved Oct 7, 2024 1 from https://www.finops.org/introduction/what-is-finops/

Fung, Bryan, "Tiktok Moves US Users' Data To Oracle Servers To Address Security Concerns," CNN Business, https://www.cnn.com/2022/06/17/tech/tiktok-user-data-oracle/index.html, June 17, 2022,

Gewirtz, David; "The cloud calculation: Factor these lock-in and switching costs into your cloud plans," https://www.zdnet.com/article/the-cloud-calculation-factor-these-lock-in-and-switching-costs-into-your-cloud-plans/; Aug 2021

Gil, Laurent, "Cloud Pricing Comparison: AWS vs. Azure vs. Google Cloud Platform in 2022." https://cast.ai/blog/cloud-pricing-comparison-aws-vs-azure-vs-google-cloud-platform/; May 2022

Glass, Erin, "A General Introduction to Cloud Computing," https://www.digitalocean.com/community/tutorials/a-general-introduction-to-cloud-computing, Nov 2020

Golightly L, Chang V, Xu QA, Gao X, Liu BS. "Adoption of cloud computing as innovation in the organization," International Journal of Engineering Business Management. Vol 14, 2022, https://journals.sagepub.com/doi/pdf/10.1177/18479790221093992

Gregalis, Jonathan. "Forced Insecurity: 64% of Adults Say Data Sharing is Required to Use Online Services Despite Declining Trust in Digital Service Providers." Businesswire, April 2022. https://www.businesswire.com/news/home/20220412005230/en/Forced-Insecurity-64-of-Adults-Say-Data-Sharing-is-Required-to-Use-Online-Services-Despite-Declining-Trust-in-Digital-Service-Providers

Griffiths, Charles. "The Latest Cloud Computing Statistics," July 2023, https://aag-it.com/the-latest-cloud-computing-statistics/

Hadidi, Jude. "Cloud Computing vs. Serverless Computing," December 2022, https://www.sitech.me/blog/cloud-computing-vs-serverless-computing#:~:text=Serverless%20Computing%20is%20a%20cloud,and%20energy%20managing%20these%20servers

Hein, David. "The 9 Best Cloud Migration Managed Services Providers of 2022," https://solutionsreview.com/cloud-platforms/the-best-cloud-migration-managed-services-providers/; April 2022

Hoffman, Jason. "6 Best Use Cases for IaaS (Infrastructure as a Service)," retrieved Feb 2023 from https://wisdomplexus.com/blogs/iaas-use-cases/

Kanade, Vijay. "What Is Platform as a Service (PaaS)? Definition, Examples, Components, and Best Practices; Feb 2022; https://www.spiceworks.com/tech/cloud/articles/what-is-platform-as-a-service/

Lardinois, F. "Google Cloud Will Now Show Its Users Their Carbon Footprint In The Cloud" Oct 12, 2021; https://techcrunch.com/2021/10/12/google-cloud-will-now-show-its-users-their-carbon-footprint-in-the-cloud/

Larkin, A. (2020, May 21). "Disadvantages of cloud computing - cloud academy blog." Retrieved April 15, 2021, from https://cloudacademy.com/blog/disadvantages-of-cloud-computing/

Peterson, Richard; "Cloud Deployment Models," https://www.guru99.com/cloud-deployment-models.html; Jan 2023.

Phalip, Julien; Niemitz, Steve. "Modernizing Twitter's ad engagement analytics platform," https://cloud.google.com/blog/products/data-analytics/modernizing-twitters-ad-engagement-analytics-platform; March 2020.

Ponemon Institute. "Global Study on Zero Trust Security For The Cloud," https://www.appgate.com/resources/ungated-resources/ponemon-global-study-on-zero-trust-security-for-the-cloud; July 2022

Ray, Chris; "The Future of ZTNA: A Convergence of Network Access Solutions," https://gigaom.com/2024/08/19/the-future-of-ztna-a-convergence-of-network-access-solutions/; Aug 2024

Robinson, Todd; "Cloud Deployment Models in 2024 – 6 Options for Modern Clouds," https://openmetal.io/resources/blog/cloud-deployment-models/ ; Sept 2024

Shapira, Ariel. "With Rising Costs and Vendor Lock-Ins, Is a Cloud Exodus in the Making?" https://www.entrepreneur.com/article/380648; Aug 2021

Shivang, 2022. "What is Autoscaling? How Does It Work In the Cloud – Simply Explained." https://scaleyourapp.com/what-is-autoscaling-how-does-it-work-in-the-cloud-simply-explained/

Tavbulatova, Z K et al. "Types of Cloud Deployment," Journal of Physics. 2020. https://iopscience.iop.org/article/10.1088/1742-6596/1582/1/012085/meta

Technavio, "Cloud Migration Services Market to witness USD 7.09 Billion growth," https://www.prnewswire.com/news-releases/cloud-migration-services-market-to-witness-usd-7-09-bn-growth--north-america-to-have-largest-share--technavio-301565827.html; June 2022

Vailshery, L, "Public cloud application infrastructure services market 2015-2024" https://www.statista.com/statistics/505248/worldwide-platform-as-a-service-revenue/ Mar 27, 2024

Vergadia, Priyanka and Prompanya, James; "How Pokémon GO scales to millions of requests? https://cloud.google.com/blog/topics/developers-practitioners/how-pok%C3%A9mon-go-scales-millions-requests; Oct 2021

Webroot; "5 Financial Benefits of Moving to the Cloud" https://www.webroot.com/us/en/resources/tips-articles/five-financial-benefits-of-moving-to-the-cloud; retrieved 10/25/2022

Wiggers, Kyle, "As Cloud Costs Rise, Startups Making Spend-Optimizing Tech See Sunny Days Ahead," https://techcrunch.com/2022/05/25/as-cloud-costs-rise-startups-making-spend-optimizing-tech-see-sunny-days-ahead ; May 2022.

Zhao, Shuyin; "Github Copilot Now Has A Better AI Model And New Capabilities," https://github.blog/2023-02-14-github-copilot-now-has-a-better-ai-model-and-new-capabilities/; Feb 14, 2023

Zhukovskiy Y, Koteleva N 2017 Method of Data storing, collection and aggregation for definition of life-cycle resources of electromechanical equipment IOP Conference Series: Earth and Environmental Science 87(3) 032057