

Public vs Private Cloud



“AI’s growth is driving tech companies toward public cloud services, offering the scalability and flexibility private clouds can’t match”

After shunning away their private physical CD distribution service in 2007, Netflix realized the potential in public online streaming. This transformation from a private service into a public network culminated in a reported revenue of \$9.4 billion within the first quarter of 2024, with over 270 million paid subscribers for the company. This type of trend has been a common disruption within many companies such as consumers paying a monthly service on Spotify instead of purchasing a vinyl record. Similarly, tech companies that originally adopted a private cloud service, which is where all their IT infrastructure internally bought and managed, are now realizing the true potential of public cloud, a more external approach to IT resources.

This transition has reached even hyperscalers - most recently, Apple announced their partnership with a public cloud service, OpenAI, to introduce ChatGPT into many Apple devices. Their negotiation with Taiwan Semiconductor Manufacturing Company (TSMC) cost Apple \$430 billion over the next five years for data center expansion and AI chip manufacturing. The race into becoming the next technological behemoth can steer the future into an ambiguous road, questioning in what ways can AI benefit or detriment mankind.



The rapid advancement of AI will provoke tech companies to shift their focus towards using public cloud data center services due to the massive amounts of data training, volume, and consumption AI requires. Unlike private cloud services that require huge capex investments, public cloud data centers are a better option as they allow tech companies to quickly adapt to the volatile AI market. This demands for greater management that a private cloud data center cannot handle, illustrating the true importance of public cloud services. As larger AI models take off, there will be such a need for external shared training data/training GPUs and other single-use equipment that outsourcing will become more common.

AI’s History of Private and Public Cloud

The rapid expansion of AI has created a greater demand for cloud services, causing business enterprises to need to outsource companies to manage their vast data infrastructure, influencing decisions on whether to use a private or public cloud service.

Cloud computing is the concept of on-demand usage of IT resources. It provides access to many different resources as and when needed. This allows users to request and release IT resources anytime, reducing the need for expensive buffer capacity. Cloud computing uses virtualization and APIs to transform hardware into remotely accessible resources. The main difference between private and public cloud is who the computing services are accessible to. A private cloud is dedicated to a single organization or entity, so no one else will have access to it. All resources are therefore isolated and there is no involvement of third-party service providers. On the other hand, public cloud services are available on-demand to anyone on the public internet. Because there are typically many other external resources involved, the range and depth of public cloud services are naturally a lot larger than private cloud services.

In a public cloud, specialized cloud providers maintain and provide the cloud service, allowing the organization to focus on its core business. The client organization has more control over the level of resources and service they use at any given time. In a private cloud, an organization owns and maintains its own data centers. Organizations with private clouds are usually large enough to justify setup and maintenance, with multiple data centers in place. It is also possible to use a hybrid cloud, in which an organization integrates internal and public cloud resources. For example, public cloud services may serve as a backup when companies need specific IT resources or exceed the private clouds' capacity. Finally, there is virtual private cloud, in which organizations retain more control over their resources and data.

According to Flexera, organizations currently use 3.4 public clouds and 3.9 private clouds on average. They aim to make use of multiple services, but ultimately find it challenging to manage costs. Cloud cost optimization was becoming a higher priority according to 52% of professionals surveyed, and 46% believed that cloud costs were on target. For companies without specialized departments, only 32% stated they were on target. To solve this problem, we predict that companies will implement cloud consolidation, bringing data into one cloud to make the most of economies of scale.

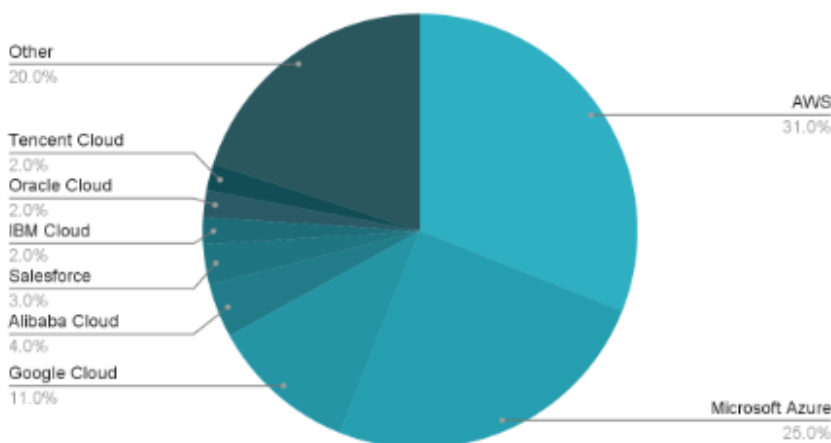
MAIN PLAYERS & CONSOLIDATION

The major players in the Public Cloud Market include Alibaba, AWS, Cisco, Google (Alphabet), HPE, IBM, Microsoft, Oracle, and Rackspace Technologies.

The major players in the Private Cloud Services market include Cisco, Dell, IBM, Microsoft, Oracle, VMWare, HPE, Rackspace, Red Hat, AWS, Google, HashiCorp, and OpenStack.

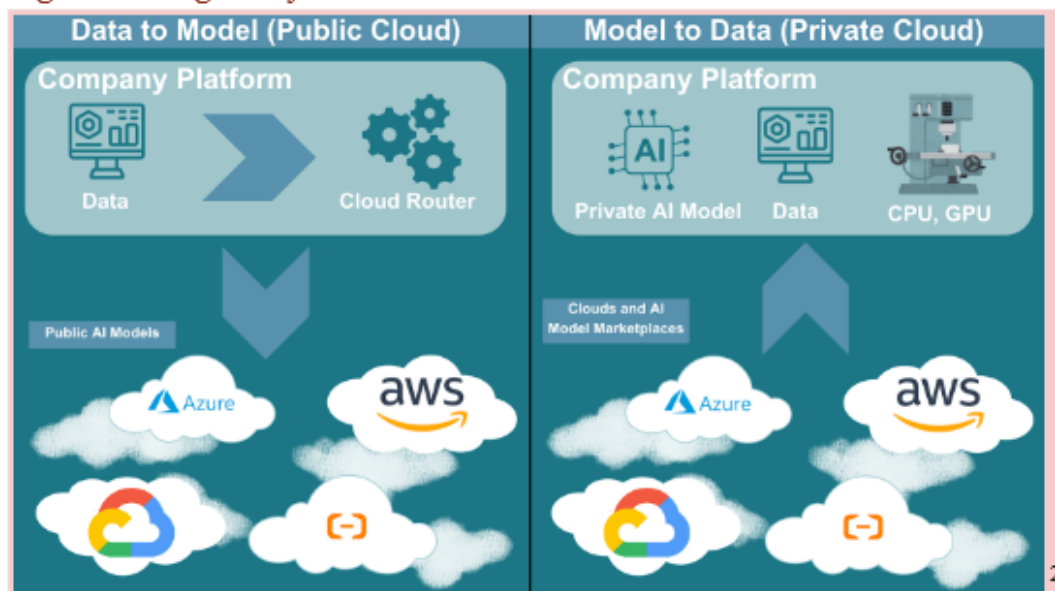
The “big three” of cloud- AWS, Azure, and Google Cloud, now dominate cloud with a 67% market share within IaaS and PaaS, translating to nearly \$200 billion in revenue as of 2023.

IaaS and PaaS Market Share



Private Cloud - At the moment, the private cloud market is valued at about USD 92.64 bn and considered a relatively high growth industry, with a forecasted CAGR of 15.8% for the next decade. The growth is mostly driven by the easy adoption and lower prices for smaller and medium sized companies as discussed later.

Figure 1: Diagram of Private and Public Clouds



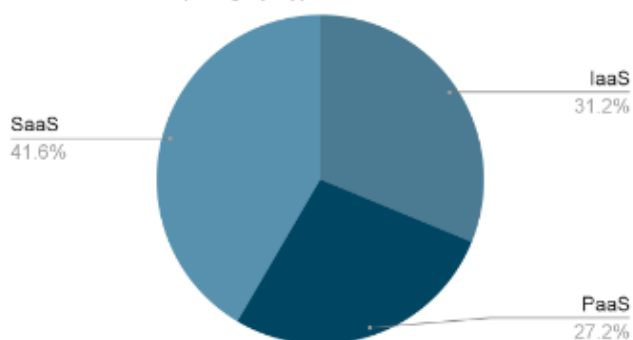
Shown in the diagram above, private cloud services allows companies to move an AI model from a public marketplace to their internal data storage to be used themselves. This is done by downloading the foundational model into their on-premise data centers, allowing internal access to private data.

This approach may be appealing to companies as it allows better data privacy, enabling direct access and security management. Moreover, companies choose their AI foundation model for their private cloud due to believing its low cost at scale. Demonstrated in Figure 3, later in the paper, the prices of private clouds range from about \$2,000 to \$34,000. Approximately 32% of enterprises use this approach as they envision that integrating AI into a private cloud service is beneficial compared to public cloud.

Public Cloud - Unsurprisingly, the public cloud market is significantly larger than the private cloud market due to how many external players can take part. The industry is currently valued at about USD 773.30 bn and expected to grow at a CAGR of about 18.5% in the next five years. Though the growth rate may not be much higher than the private cloud industry, due to the sheer size difference in the industries it is much more remarkable how quickly the public cloud industry is continuing to grow.

Most of this computing is for modern AI spending. Of this spending, \$156.2 billion on infrastructure as a service (IaaS), \$136.4 billion on product as a service (PaaS), and \$208 billion was spent on software as a service (SaaS). In general, infrastructure as a service allows clients to directly manage underlying infrastructure but scale more efficiently. Platform as a service involves providing outsourced IT infrastructure and a development environment, often also reducing coding needs. Software as a service involves providing an application that is fully formed but with some centralized computing, such as Google Drive or Salesforce. We have provided more details and examples for these subsectors in Appendix B. It is important to note that SaaS applications are often built on PaaS platforms, which themselves are often built on IaaS platforms. A large plurality of cloud computing solutions are SaaS because the end customer prefers a vertically-integrated solution.

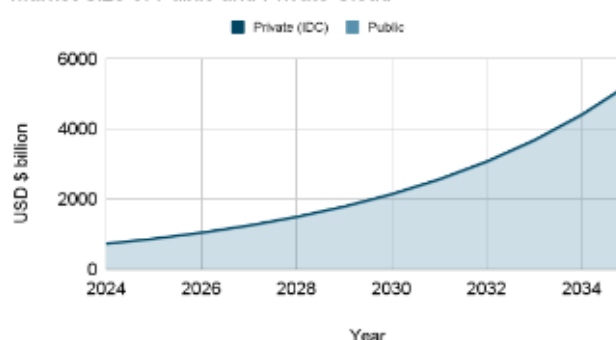
Public Cloud Computing by Type



About 32% of companies transfer their private data infrastructure into a fully-managed AI cloudspace, or a public cloud. Companies may choose to adopt this approach due to the cloud's low barriers of entry and due to the plethora of resources it provides. Advanced GPU and CPU models that are only available in cloud marketplaces provide the opportunity for a business to get their high quality AI initiative up and running instantly, whereas private clouds require internal management that can hinder their position in the AI arms race.

More specifically, public cloud users often get away in putting effort to upgrade and test a private cloud AI model, which leaves the IT staff to focus more on other important tasks. In addition, the unprecedented expansion of AI will reduce the number of available IT engineers, especially AI engineers. Those who pursue the private cloud approach would need to continuously demand for AI engineers in order to fully upgrade their AI model. Therefore, companies need to pursue a public cloud service due to the current shortage of AI talent within their company.

Market Size of Public and Private Cloud



General overview between the two clouds:

Public Cloud (Outsourced)	Private Cloud (Internal)
Third-party manages data centers and technical details	The company manages its own data centers and technical configuration
Broad, high-quality expertise – technology tends to be “develop once, use repeatedly”	Limited offerings due to lack of economies of scale
Simple deployment using API/GUI	Requires in-depth understanding of underlying technologies
Spread out operating expenses	Upfront capex and high maintenance costs
Better AI models available at much lower cost ⁴²	Data security, model monitoring, and nimbler models ⁴³

Others may argue that the private cloud approach may be appealing due to better autonomy, security, and cost. However, it is important to note that private clouds purely impose responsibilities onto companies that have no AI expertise, which would slow them down in keeping up with the AI arms race. Civo, a company that specializes in offering cloud services, studied that over 50% of businesses across various industries lack dedicated machine learning workers.

This further emphasizes the need for public cloud services as private cloud services require companies to manage their own AI foundation model, with or without AI expertise. Hence, within the next twelve months, 54% of companies with 1000+ employees are moving their data into the public cloud network. This piece of information illuminates the strong importance of moving internal data infrastructure to public cloud services within an environment lacking AI engineers.

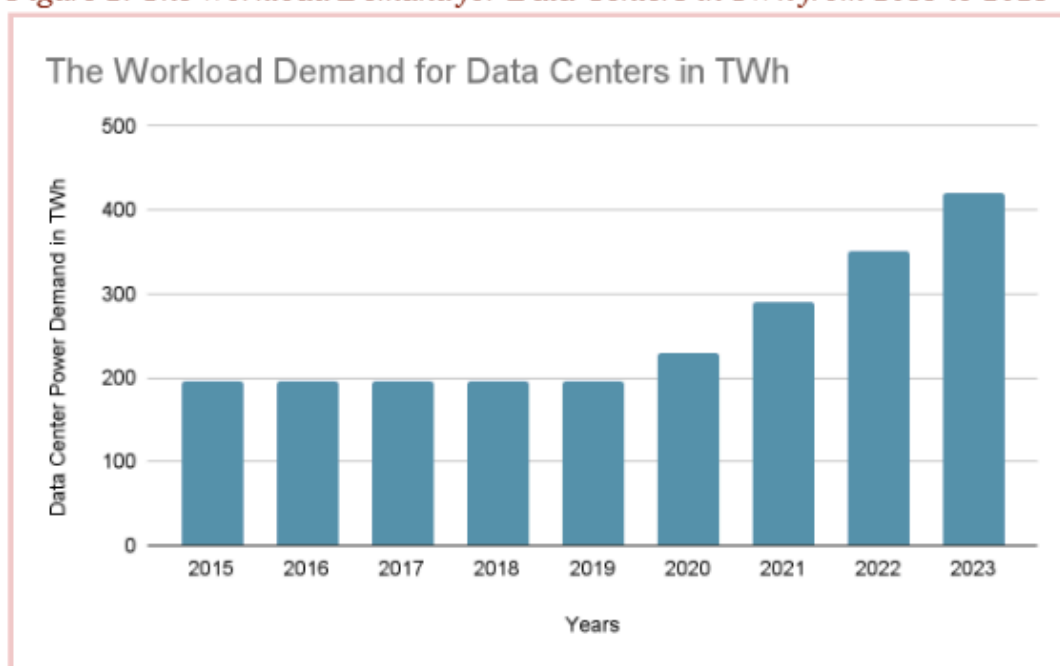
Although it may be dubious to upload one's private data onto a public cloud network, companies can still keep up with the AI arms race as they utilize high-quality models that are constantly updated with the latest version in the cloud market space. These types of models are only available within the public cloud network, making private cloud a distant memory for businesses.

Evidence indicates that public cloud services should be the optimal approach in the upcoming years given the shortage of AI engineers and the cloudspace's availability of efficient models. Because of this, we expect the public cloud industry to continue growing at much faster rates than the private cloud industry.

Data Center Expansion and its Effect on Private and Public Clouds

Much of this shift to public clouds has also been driven by the expansion of data center throughout the world. Along with cloud networks, Goldman Sachs Research estimates that data center power demand will grow 160% by 2030. Thus, many hyperscalers have been scrambling to expand their data center base and form partnerships with AI companies in an attempt to lead the race. For example, Microsoft has recently announced a plan to collaborate with NVIDIA to create a supercomputer data center over the next five years. They invested roughly \$100 billion into Open AI to create Stargate, the AI-centric data center that would be available within the public cloud market in 2028. Illustrated in Figure 2, the introduction of AI in 2021 has boosted the need for data centers as companies need a greater amount of energy to power AI operations.

Figure 2: The Workload Demand for Data Centers in TWh from 2015 to 2023



Environmental Factors of Public vs. Private Clouds - The growing demand for data centers from private and public cloud providers poses significant environmental risks, with social costs from pollution estimated at \$125–\$140 billion. Hyperscalers, like Google, are major contributors, as AI models such as AlphaGo Zero produced 96 tonnes of CO₂ in just 40 days—equivalent to 1,000 hours of air travel. With rising emission fees, tech firms are expected to invest heavily in renewable energy to meet ESG goals, as 75% of business leaders plan to enhance their sustainability strategies. Energy efficiency will be crucial when choosing between private and public cloud services.

While private clouds can be seen to be cheaper, the emission cost for private clouds tells otherwise. If a company were able to use AI such as GPT-3 within a private cloud, they are expected to confront a cost of 1,300 megawatt hours of electricity that can power 130 US homes annually. Depending on the region, one megawatt per hour costs between \$24 to \$46, totaling an estimated production cost of \$31,200 to \$59,800. Furthermore, implementing AI into private cloud networks would actually require about six to 24 months to train, taking up a company's resources and efforts that can be spent in other areas. This would continuously add up as companies will forgo the opportunity of having a highly efficient AI model only available in the public cloud network, slowing them down in the AI arms race.

Meanwhile, those companies who utilize public cloud softwares are continuously using updated versions made by public cloud providers such as AWS, Google Cloud, and Cisco. Unlike private clouds, the public cloud model is much more energy efficient, where AWS' infrastructure is 3.6 times more energy efficient than a typical US data enterprise center. Since the public cloud approach aligns more easily with emission regulations, it would reduce long-term operating expenses. This piece of information further highlights how social pressures due to environmental effects will continue to push the public cloud industry further as ESG becomes an increasingly relevant topic.

Financial Benefits Behind Public Cloud

However, the environmental benefits public clouds provide do seem to come with a steep cost. As shown below, public cloud tends to be between 4-6x more expensive than private cloud.

Figure 3: General Costs of Public and Private Clouds

Deployment Size, Bandwidth	Public Cloud	Private Cloud	Difference in Cost
Small - 100VMs, 10TB	\$7,731	\$1,952	\$5,779
Medium - 500VMs, 50TB	\$28,925	\$7,375	\$21,550
Large - 1000VMs, 150TB	\$60,450	\$11,533	\$48,917
XL - 2000VMs, 300TB	\$86,260	\$18,570	\$67,690
XXL - 3000VMs, 600TB	\$123,061	\$25,770	\$97,291
XXL - 4000VMs, 1200TB	\$172,571	\$33,742	\$138,799

Given that private clouds are cheaper as companies do not have to pay operational expenses to the public cloud network, choosing private over public cloud makes logical sense. However, other sources support how it is still costly for businesses to implement their AI foundational model into a private cloud service as they have to manage it entirely by themselves. Even for a small business, an AI software development team can cost up to \$320,000 per year. Moreover, estimates show that the continuous maintenance of the AI model can roughly cost a private-cloud owned business \$60,000 per year. These private cloud service costs can stack up over time as a company's research and development team has to continuously upgrade their private cloud AI model.

This differs from a public cloud service as users will only have to pay for consumption, not the development and maintenance of an AI model. Hence, modern businesses are starting to recognize the secret value behind public cloud infrastructure as migrating data into a public cloud infrastructure can actually reduce the TCO up to 40%.

Overall, while it may seem that private cloud services are a cheaper alternative, it is important to note for companies that these costs can stack up over time towards \$380,000 per year. Public clouds are a better long-term investment as not only do companies only pay for consumption, but also they are keeping up with the AI arms race as they utilize a highly proficient model daily. Therefore, while we may see a short-term influx in private cloud due to the costliness of public cloud, the long-term benefits of no upkeep costs of public cloud will help the industry continue to have sustainable growth.

The Rise of External (Public) Cloud Servers - Hyperscalers' Integration of AI

Recognizing the surge in demand for AI, major cloud providers have raced to establish strong partnerships with AI companies like OpenAI and Anthropic. So far, Microsoft has invested over \$13 billion – or 13x 2024 revenue in OpenAI. Additionally, it has served as the only provider of cloud compute for all OpenAI activities, from training to APIs, and has also applied OpenAI technologies in search, marketing, code suggestions, and even Microsoft Office.

Anthropic, a major OpenAI rival, has also received significant investments from cloud services companies, signaling their transition to the public space. The company, founded by former OpenAI employees, aims to enhance AI safety. Last October, Google agreed to invest \$2 billion in Anthropic. More recently, Amazon invested \$4 billion (\$1.25 billion in September 2023 and \$2.75 billion in 2024) – this nearly matches the company's 2023 valuation. In return, Anthropic will primarily use AWS for cloud computing and make its foundation models available on Amazon Bedrock. Additionally, AWS, Anthropic, and Accenture have partnered to facilitate custom and responsible generative AI products. Such mutual agreements establish near-exclusive relationships between specific cloud computing giants and specific AI companies. This guarantees a steady stream of revenue for data storage and computing power. In contrast, other companies invested much less in the area, with Salesforce investing only \$250 million in Anthropic and Zoom investing an undisclosed amount, so they will not be able to dominate the relationships.

For all of their spending to date, Alphabet, Amazon, and Microsoft still have plenty of funds left, with \$177 billion of cash, cash equivalents, and ST investments currently on hand.

Additionally, technology spending tends to have high operating leverage. It has high fixed costs: a typical server costs \$30,000, and the price for cutting-edge GPUs will soon exceed \$70,000 per chip. In contrast, variable costs are low despite the high energy usage since electricity costs so little. The cost to power one 700W-GPU is only \$2.94 per month at a rate of \$0.13/kWh. With such economies of scale, it is impractical to expect every organization to purchase its own AI cloud infrastructure. In fact, organizations can save hardware costs, such as compute, storage, and networking, by up to 40 – 66% when migrating to public cloud. Therefore, it makes sense for them to continue investing in public cloud infrastructure.

As they establish more partnerships and provide convenient development environments, we see these companies expanding their dominance. We project that by 2035, these three public cloud providers will gain nearly 90+% market share (\$4.7 trillion+) across IaaS, PaaS, and SaaS, especially in advanced applications.

AWS	     
Microsoft Azure	  
Google Cloud	  

Why the Advantages of Private Cloud are Illusory

Not only does public cloud offer many advantages, but it also seems as if the benefits of private cloud are not as favorable as they seem. Considering the fact that the main advantage of private cloud services is the security it offers, one would assume there would be significantly less security breaches in private cloud servers. However, there have been many security breaches within both private and public cloud services, amplifying the possible dangers once AI is infused within these two systems. As of 2024, 45% of data breaches occurred in the cloud, which cost an average of \$4.59 million. With such hefty financial consequences, it makes sense how the global cloud security software market was sized at \$34.3 billion in 2024 and was growing at 24.7%. Shockingly, only 11% of companies have encrypted over 80% of their sensitive data in the cloud, so there is still space for progress. Here are some examples of private and public clouds that experienced security breaches:

Cloud Product	Cloud Type	Incident Description
Cisco	Private	- April 2024, hackers subvert digital security devices to break into national government network⁹² - Target on routers and technological devices are popular, reside in the data infrastructure perimeter⁹³
AWS	Private/Public	- Pegasus Airlines, an AWS client, had 23 million personal files leaked⁹⁴ 6.5 terabyte AWS S3 Bucket, public cloud storage, was unprotected⁹⁵
IBM	Private	- Johnson & Johnson's Jassen CarePath patient support program was leaked⁹⁶ 1.16 million patients' private data was stolen⁹⁷

Although security breaches are always inevitable for both private and public cloud systems, some may argue that private cloud services should be preferred from this framework due to higher autonomy and management of personal data. Enterprises have easier access to data, which can be further emphasized when using AI within its own data infrastructure. Another common claim is that it is often challenging to enforce privacy guarantees from providers and limit privileged access.

However, it is important to note that public cloud services are managed by professionals, especially those who have experience in designing IT. Instead of a company individually managing its own private data with no IT experience whatsoever, the public cloud is actually typically more secure as it is controlled by individuals who understand the difficulties surrounding security breaches.



Cybersecurity in the public cloud also benefits from economies of scale. Public cloud providers can fund security upgrades, especially software upgrades, with funds from many more clients.

Additionally, public cloud providers have a reputation that allows them to hire the best IT professionals. Large cloud providers also have an advantage regarding scalable security measures such as physical fortification and multi-factor identification.

Also, companies who choose private cloud often suffer from “perimeter complacency.” Private cloud is much less secure than it seems, as employees could inadvertently download malware over the internet. The main factor protecting a private cloud is its “obscurity”: as companies become more well-known their data is less secure without professional protection.



Legal software providers such as Thomson Reuters 3E and FINRA, the primary regulatory organization of broker-dealers, have chosen public cloud over private cloud due to its many advantages.

For those companies that are still worried about cloud security, there are a number of specialized service providers such as Palo Alto Networks, CrowdStrike, and Trend Micro. These companies are more willing to work with large cloud providers, often through the marketplace, with over \$1 billion of CrowdStrike’s \$3.3 billion revenue coming from AWS alone.

With all the external cybersecurity resources out there, such as the previously mentioned legal software providers and other specialized services, public cloud servers actually have the upper hand with cybersecurity, thus taking away private cloud server’s main pro.

Conclusion

As artificial intelligence launches us into a new age, companies face a dilemma crucial to their survival - should they continue focusing on private cloud servers or devote more toward public servers? On one hand, the insights and efficiency that come with automated analysis and the external resources that public cloud servers would bring provide companies with an indispensable competitive advantage. On the other hand, it could help to keep the data that power these world-changing new technologies confidential through private cloud servers to maintain advantages. However, due to the rapid pace that technology is evolving with combined resources and the lacking advantages of private cloud servers, there is a strong chance more and more companies will transition to public cloud.

While private clouds may seem appealing due to its supposed increased security and lower price, companies also must be aware of the amount of time and effort they must put to maintain their AI model against other competitors. Just hiring an AI development team to maintain an internal private cloud AI model alone would cost up to \$380,000 per year. The burdensome investment into the R&D for AI would continuously add up, bringing more weight to the shoulders of those who utilize private cloud resources. Hence, many companies have been moving their data into public cloud networks, where studies demonstrate that migrating to a public cloud network would actually boost profit growth by 11.2%. It is undeniable that public cloud will be a stronger investment in the long-term than private cloud.

Additionally, as previously seen, any security benefits are illusory as private clouds also access the internet. Private clouds will be reduced to a limited range of highly-sensitive applications in financial services, healthcare and pharmaceuticals, and the legal system. Even in these areas, the largest companies will likely pursue hybrid cloud or public cloud. Only the most sensitive workflows (often those processing original data) at smaller firms will stay in private clouds.

Lastly, large public cloud providers are at a significantly advantage in the AI race through partnering with professional AI developers, further giving a great incentive for many enterprises to utilize the public cloud. Microsoft's recent multi-year partnership that resulted in \$100 billion with NVIDIA makes it one of the only public cloud networks that utilizes NVIDIA's GPU, networking and software. This high-tech availability for public cloud users allows them to keep up with the rampant AI arms race, a lacking characteristic that private clouds have.

In conclusion, using the public cloud provides economies of scale from functionality, security, and energy efficiency perspectives. Public cloud is like a highly-optimized reservoir, while a private cloud is like getting drinking water from a pond. Just as it is impractical and unsafe to get "cheap" drinking water in the wild, it is utterly impractical to use the private cloud in an age of advanced technology. Therefore, private clouds will be obsolete by 2035 despite predictions to the contrary. As companies require additional access to advanced technologies like AI, they are starting to recognize the true potential of large public cloud services as users grew at 20.4% from 2023.