

# Which Infrastructure / OS for Our Next Startup?

SOOBIN RHO<sup>1,\*</sup>

<sup>1</sup>Technical Co-Founder of PermaThreads

\*soobin@permathreads.com

Compiled May 12, 2023

First of all, the purpose of this paper is to answer the question, “Which infrastructure / OS for our next startup?” From the perspective of a start-up founder, this paper surveys on (a) cloud servers vs on-premise servers, (b) auto-scaling cloud servers vs dedicated servers, and (c) which operating system.

<http://github.com/PermaThreads>

## 1. INTRODUCTION

For context, the author and four of his classmates (hereafter, we) are participating in 2023 Hult Prize, which is like Call for Code, but more focused on creating a social enterprise. In it, we are trying to create a start-up that is beneficial for both the planet and people, and at the same time with good profitability. PermaThreads is a platform that extends the wear life of clothing and promotes cyclical fashion by connecting users with tailors/entrepreneurs in their area. This paper addresses the question, “Which infrastructure / OS for our next startup?”

## 2. CLOUD SERVERS VS ON-PREMISE SERVERS

We first had to choose between cloud servers (off-premise servers) and on-premise servers. Long story short, we decided on cloud services. The next section of this paper discusses why we decided provisioning cloud services is more adequate than getting an on-premise server for our use cases for now.

On-premise servers refer to computer systems that are physically present in our own facility, whether it be a part of our office or our own data center. On the other hand, cloud servers are computer systems that are hosted by cloud service providers, such as Amazon Web Services (AWS), Google Cloud, and Microsoft Azure.

The main difference, which by the way also happens to be our biggest deciding factor, is the initial set-up cost. (Villamizar et al., 2016, pg. 1) Getting an on-premise server requires purchase of all the computer hardware that are required for the server specifications. Building a server with consumer-grade hardware can cost less, ranging from \$30 to thousands of dollars but building a server with specialized hardware specifically designed for servers, equipped with greater hardware reliability, can cost millions of dollars.

Of course, our start-up doesn’t need server-specific special hardware because that extra reliability that comes with those special server hardware is not marginal enough for start-ups

like us. We would instead use consumer grade hardware that cost much less, looking at a few thousands of dollars, but even then, this initial set-up cost is significantly higher compared to the initial set-up cost of getting a cloud server. For example, many of the cloud servers adopt the pay-as-you-go model, which means the initial set-up cost of a cloud server can be as low as \$10 for the first month. Therefore, we decided to use a cloud server when we first launch our startup.

However, we plan to at least consider getting an on-premise server in the future because we expect there to be a point in the far future, where getting an on-premise server would actually cost less in the long term compared to using a cloud server. Yes, the initial cost of on-premise servers is greater than that of cloud servers. Nonetheless, as the startup grows and the server usage goes up, the cost of using a cloud server can accumulate and actually cost more in the long term compared to using an on-premise server.

In addition to the set-up cost, there are several differences between on-premise servers and cloud servers. Setting up an on-premise server requires system designs of which hardware to use depending on the server use cases; shipment of said hardware, which usually takes days; assembly of these hardware; OS installation; and continual system updates and security patches. Furthermore, getting an on-premise server requires server-room maintenance with A/C units for proper temperature control, and there also needs to be proper electrical-power management to ensure there’s enough power supply for every part of the system.

If any of these steps are not initially set up properly, or if they go out of date due to negligence or various reasons, serious consequences follow. For example, failing to update software regularly would introduce security holes and vulnerabilities in the system, which would then allow malicious actors to not only control the system, but also leak sensitive information, such as users’ personal information. This can not only violate the privacy of users and ensue a number of legal actions, but also endanger lives, if those personal information also contains medical information. Also, inadequate set-up of electrical systems of the server room can result in system failures whenever there’s a surge in energy consumption. This would mean down time.

Unlike on-premise servers that require physical space in our own facility, cloud servers are, by their own definition, managed by cloud service providers. Since the management of the infrastructure is already handled by them, provisioning a cloud server takes much less time than getting an on-premise server – some of them being as fast as just a few seconds. There’s no need to

allocate a physical space on our own facility. Nor is there any need to worry about server maintenance, since it's the cloud service providers' responsibility to maintain the server. (Namasudra, 2018, pg. 2) Cloud servers offer "reliability ... [and] resiliency." (Namasudra, 2018, pg. 2) In other words, cloud services allow start-ups to start developing "without carrying on costs and risks of building and managing their own hardware and infrastructure. [For example,] many successful companies like Dropbox, Zynga and Netflix to name a few, successfully rely on AWS." (Bermudez et al., 2013, pg. 1)

At the same time, cloud servers have downsides: "security and privacy in the cloud, transferability, downtime, understanding and limited control." (Namasudra, 2018, pg. 2) "If the cloud provider stops their services due to a financial crisis or due to some other reasons, there may be data loss. Users are not able to access data in future because data are not available on the cloud servers. So, data loss is a major problem in cloud computing." (Namasudra, 2018, pg. 10) For example, OVH Cloud had a fire at one of its data centers in France, and this resulted in many of the websites hosted in OVH Cloud to go down and lose data permanently. (AP News, 2023, Fire at French cloud computing firm disrupts websites)

Plus, "cloud computing has many risks for maintaining the confidentiality of data ... if user's data is stored outside their country, the government of that country can view the actual data under the circumstances ... Another main issue in the cloud server is that how can users know that the data or file are fully deleted from the server, and it cannot be recovered again. All the data on the cloud server can be backed up by the CSP [Cloud Service Providers]. Currently, there are no procedures to know that user's data is completely deleted from the cloud server or not. The CSP stores data in several physical devices. In one device, there are many data, and the device cannot be destroyed from the cloud server. Therefore, it creates a critical data issue for DOs. [Data Owners]" (Namasudra, 2018, pg. 10) Therefore, when choosing a cloud server, it's crucial for us to consider the country the data will be hosted in. "For the protection of user's confidential data or sensitive data, legal frameworks are a major key." (Namasudra, 2018, pg. 15)

Despite these downsides, however, we are not trying to discourage anyone from using a cloud server. Instead, what we aim to do here is to be informed about both the strengths and weaknesses of whichever infrastructure we choose to implement. In fact, to answer the question, "cloud servers vs on-premise servers," we choose to provision a cloud server at the beginning of our product launch, and continue to do so until there's a need to do otherwise. Namely, if the startup grows and the server usage cost from the cloud server becomes high enough to justify getting an on-prem server instead as a long-term investment, we will consider migrating to an on-premise server. Of course, this is not a final decision, meaning that we will definitely move away from cloud servers for sure when it becomes too expensive to do so; rather, it's just that those kinds of possibilities will open up at that stage of our start-up where we can consider these options. Moreover, it might be the case that staying on the cloud server proves more effective in terms of saving money - and most importantly - serving our users. There might even be unforeseen factors that will make us either stay or move away from using cloud servers or on-premise servers. Thus, our answer to the question, "cloud servers vs on-premise servers," is yes - both are adequate depending on our use cases. We will start with a cloud server, but we will be open to both as our needs evolve, and we will adapt and make a decision accordingly.

### 3. AUTO-SCALING CLOUD SERVERS VS DEDICATED SERVERS

Now that we decided on using a cloud server instead of an on-premise server, the next thing we had to decide on was which cloud servers to use. Again, long story short, we decided to use auto-scaling cloud services. Before diving into the main discussion of auto-scaling cloud services vs dedicated servers, this section first provides context on what kind of services are provided as cloud services.

"Each consumer can opt either to share access to the software with other consumers (multi-tenancy), thus enabling shared total costs and creating economies of scale, or decide to be a single tenant, thus providing greater control and security." (Goyal, 2014, pg. 3) Single-tenant cloud services, such as "baremetal is a non-virtualized physical server ... Because baremetal is a dedicated server, flexibility and performance are high but provisioning and start-up time are long." (Yamato, 2018, pg. 4) "The idea of bare metal cloud servers is to give the end users full processing power of the physical server without using a virtual layer ... Since the server is not shared among multiple tenants, no one can interfere with the performance of the machine. This technology removes the overheads of virtualization, while providing the high availability and elasticity of the cloud." (Rad et al., 2015, pg. 1) In contrast, multi-tenant cloud services tend to be less powerful than dedicated servers, but tend to take a shorter time to provision. "Compute resources are isolated with each unit called a container but the OS kernel is shared among all containers. Containers do not have kernel flexibility, but container creation only requires a process invocation and a short time for start up." (Yamato, 2018, pg. 4)

As a side note, there are security concerns related to virtualization in cloud services. "For using virtualization, new security issues like cross-VM side-channel attacks are occurred." (Namasudra, 2018, pg. 14) Especially, "when virtual machines are migrated from one place to another, this attack [migration attack] occurs. Many IT companies move their virtual machines to various places according to their uses. Attackers attack those machines, and hack user's sensitive data."

To us, the question, "auto-scaling cloud services vs dedicated servers" boiled down to costs. As a start-up, it's detrimental to minimize the server cost in order to survive. We decided to go with auto-scaling cloud services because of their pricing model - i.e. pay-as-you-go model. This is in contrast to the model that is adopted by dedicated servers, where the server cost is by the minute. In auto-scaling cloud services, we only pay for the system resources usage and the web traffic coming in and out of the server. "The use of emerging cloud services such as AWS Lambda, exclusively designed to deploy microservices at a more granular level (per HTTP request/function), allows companies to reduce their infrastructure costs in up to 77.08%." (Villamizar et al., 2016, pg. 5) Also, response time is a metric that is important in many cases too, but in our case, where we are trying to be sustainable in the long term and to survive as a company, we decided that it's safe to assume cost is more important than response time for now at the beginning of our product launch. Of course, being too slow will harm the user experience, so we definitely will minimize the response time, but that will come only after our start-up stabilizes.

### 4. WHICH OPERATING SYSTEM

Again, long story short, we decided to use Ubuntu Server LTS (Long-Term Support). We considered Unix OS's, such as

FreeBSD, and other Linux OS's, such as Debian and Fedora, as well as Windows Server, but the main deciding factor here was the compatibility with the tech stack we are using – Docker and Kubernetes + HTML, TailwindCSS, TypeScript, and Next.js. Ubuntu Server not only has the support of all of the components in our tech stack, but it also had the biggest community support, by which we mean amount of documentation + StackOverflow answers.

Server Types. *J Netw Syst Manage* 26, 339–360 (2018). <https://doi.org/10.1007/s10922-017-9418-z>

AP News. (2023). Fire at French cloud computing firm disrupts websites. [online] Available at: <https://apnews.com/article/paris-fires-cloud-computing-france-412f64e9807fc94d8dca4412536a4899> [Accessed 23 Apr. 2023]

## 5. 2 YEAR ROADMAP

Each specific start-up has different needs, so keep in mind that the findings of this paper may not apply to other start-ups in the same way. In the same vein, keep in mind that new technologies keep getting invented. What this means to us, founders of start-ups, is that we should always be on the lookout for the possibility of finding a better way of doing anything no matter what stage our product is at. It is harmful to just seek out the newest technology just for the sake of using new technology, of course, but it is essential to be able to migrate when necessary as a company. Here's our roadmap for the next two years:

Apply to AWS Activate Founders, Google for Startups Cloud Program, and Microsoft for Startups Founders Hub. → See which one we get into. They offer one year of free credit ranging from \$1,000 to \$100,000 on average. If we get accepted by all three, use all of them in turn. Try them out and see which one best suits our needs and purposes. → Calculate if it's more effective to stay at a cloud server or to get an on-premise server outright.

## 6. REFERENCES

Rad, P., Chronopoulos, A.T., Lama, P., Madduri, P. and Loader, C. (2015). Benchmarking Bare Metal Cloud Servers for HPC Applications. 2015 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM). doi:<https://doi.org/10.1109/ccem.2015.13>.

Mukherjee, S. (2019). Benefits of AWS in Modern Cloud. SSRN Electronic Journal. doi:<https://doi.org/10.2139/ssrn.3415956>.  
Namasudra, S. (2018). Cloud computing: A new era. *Journal of Fundamental and Applied Sciences*, [online] 10(2). doi:<https://doi.org/10.4314/jfas.v10i2.9>.

Abidi, F. and Singh, V. (2013). Cloud servers vs. dedicated servers — A survey. 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE). doi:<https://doi.org/10.1109/mite.2013.6756294>.

Villamizar, M., Garces, O., Ochoa, L., Castro, H., Salamanca, L., Verano, M., Casallas, R., Gil, S., Valencia, C., Zambrano, A. and Lang, M. (2016). Infrastructure Cost Comparison of Running Web Applications in the Cloud Using AWS Lambda and Monolithic and Microservice Architectures. 2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). [online] doi:<https://doi.org/10.1109/ccgrid.2016.37>.

Sumit Goyal, "Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review", *IJCNIS*, vol.6, no.3, pp.20-29, 2014. DOI: <https://doi.org/10.5815/ijcnis.2014.03.03>

Yamato, Y. Server Selection, Configuration and Reconfiguration Technology for IaaS Cloud with Multiple