Introduction to Cloud Computing - Part 1

This lesson provides an introduction to cloud computing.

> **We'll cover the following** ^
>
> - What is cloud computing?
> - Hosting on the cloud: The way forward
> - Cloud service providers

What is cloud computing?

Cloud computing, commonly referred to as the *cloud*, is the delivery of computing power over the web. This means the cloud service providers provide the infrastructure to the businesses over the web via their user-friendly dashboards and *APIs*, enabling them to host their services without the need for setting up any sort of infrastructure on their premises.

The infrastructure that is made available to these businesses is typically web servers to run their apps, disk storage to store massive amounts of data, compute-heavy servers for rendering animation, run analytics on *Big data*, run computational heavy scientific algorithms and so on.

All the information exchanges between the business and the cloud provider happen over the web in an encrypted fashion via an *API*. We currently live in an API driven world. Everything is API driven and is available as a service powered by an API. There is even a term for this, called *Everything as a Service*.

In today's tech landscape, there is a surge of *as a Service* offerings like *Analytics as a Service, Blockchain as a Service, AI as a Service, Games as a Service, Payments as a Service, Monitoring as a Service, Notifications as a Service*, and so on.

Today, most of the world is online, and the part that isn't will eventually be online. And as the consumers move online, the businesses follow. There is a surge in the number of web-based startups emerging in every niche, including mobile gaming, e-commerce, daily commutes, and others.

As the number of startups increases, there is an ever-growing need for the infrastructure they can use to run their services. This requirement is fulfilled by cloud service providers. The gist is that, in today's tech landscape, the cloud business is booming and is expected to continue to grow at a rapid pace.
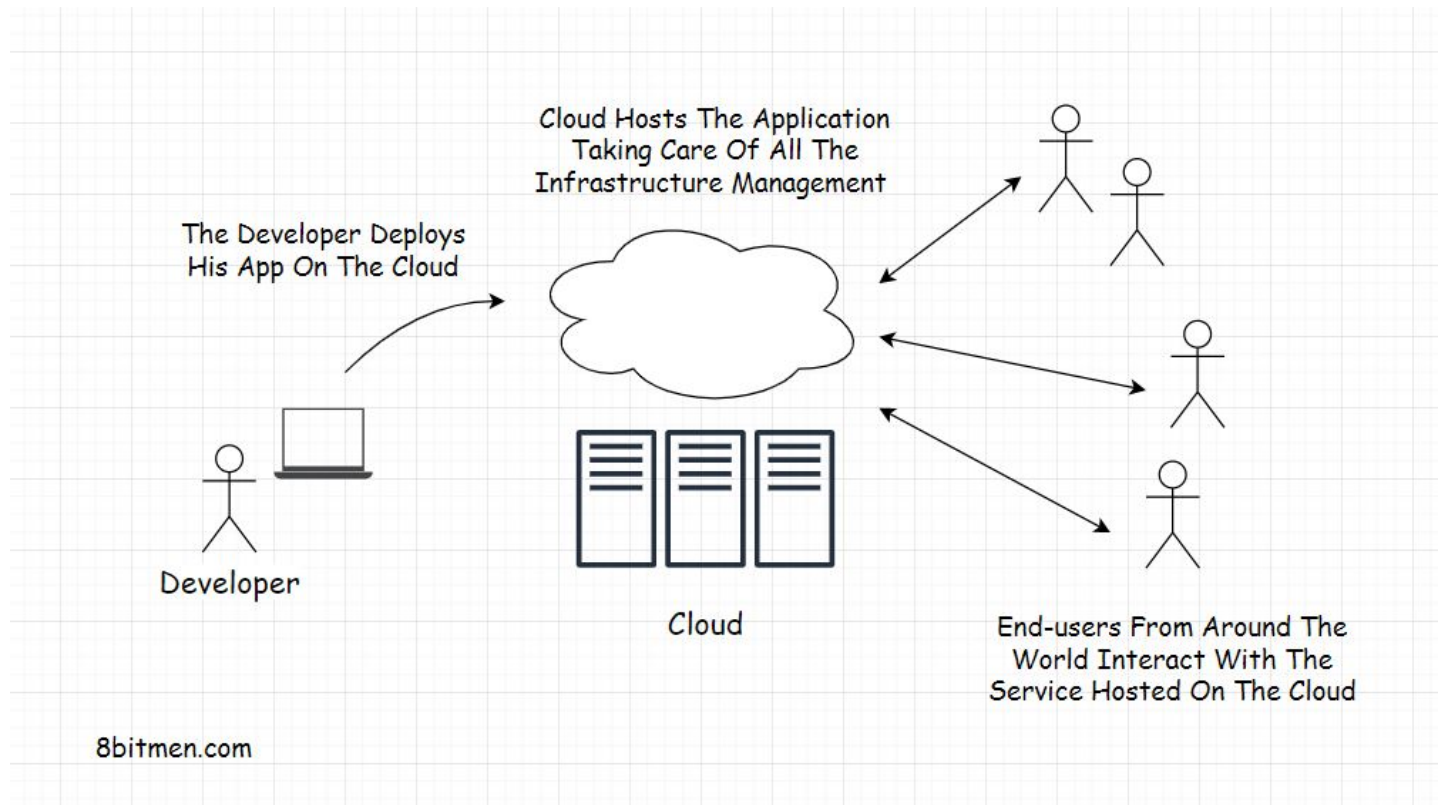
Hosting on the cloud: The way forward

In the present state of the software development universe, I am yet to come across a developer who, after developing their app, says, *"Hey, I've written this cool app. Now, I have to rush to the computer store, buy some bare metal servers, install Linux, and get them up and running. I guess I should set up everything in my basement. Right?"*

Ideally, the next move after developing an application is finding the right platform to host the app. We browse through the web looking for the right cloud platform that fulfills our infrastructure requirements and, of course, fits our budget.

On the contrary, we sometimes even pick the platform or the cloud product first and then build our app using the platform's APIs. This generally happens when using a *PaaS (Platform as a Service)* to run our service. PaaS is a cloud service model; we will discuss it in detail in the upcoming lessons.

Cloud service providers

So, as the cloud service industry expands, the cloud service providers continue to onboard more and more businesses, hosting both startups and big enterprises simultaneously.

These businesses prefer to build their products from the bare bones to run on the cloud. Additionally, those that are running legacy workloads on-premises are either planning or in the process of migrating to the cloud with the technical assistance of the service providers. We will talk about what it means to be on-premises, or on-prem, in a later lesson where you'll be introduced to different cloud deployment models.

Speaking of the big players in the cloud service space, like the *AWS, Google Cloud Platform (GCP), Microsoft Azure*, have set up massive, highly resilient data centers across the globe over time. They can scale with any kind of load pretty quickly using the plug and play commodity server approach. Over the years, these big companies have mastered the art of running and scaling the cloud infrastructure with minimum downtime.

Here are a few popular services that are scaled with millions of users and deployed on cloud service providers:

- *Pokémon Go*, the augmented reality mobile game that became a phenomenon, was developed by *Niantic* and hosted on the *Google Cloud Platform*. According to the Google Cloud blog, right after the launch, the game attracted a massive number of players up to *fifty* times the anticipated traffic, and it had to scale exponentially to ensure a smooth gaming experience to the players.

- *Fortnite* by *Epic Games* is an online multiplayer game that runs on *AWS* servers that handle a load of a whopping *two-hundred* million players (according to this case study). *Fun Fact: two-hundred* million is *twenty-four* times the population of *New York City* .

- *Hotstar*, a leading streaming media and video-on-demand service in *India* with a user base of over *two-hundred* million users, runs its services on the *AWS* cloud. In the latest edition of the *IPL T20* Cricket tournament, the streaming platform received record traffic of *10.3* million concurrent users smashing its previous record of *8.26* million concurrent users. Here is a post on it that I've written on my blog.

These cloud providers have achieved *economies of scale*. They run the services of so many businesses around the world in their data centers that collectively the cost of running a service for a single customer comes down significantly in comparison to the costs associated if the business self-hosts its service *on-premises*.

Just in case, you are intrigued about how data centers operate, here is a video of Google giving an internal tour of their data center on YouTube.

Let's carry this discussion forward in the next lesson, where I talk a bit more about the cloud service providers, the history of cloud, and how it all started.