

Serverless - Heikki Pulli

1

Serverless has many definitions because of the different ways to look into it. It is a term meaning of hiding the server management from app developers and therefore it also means of thinking apps just as functions that provide something valuable. To me all computations are serverless as long as only thing I need to know is how much does it cost to run my app on providers platform and that I don't need to pay from the time nobody is using my app.

2

Probably Faas (Function-as-a-service) because to me apps are not just some functions that do stuff. Apps consists of multiple different functionalities that are managed to work together. These single functionalities can be abstracted into functions, but alone they don't consist of working app.

3

FaaS is a single function that provides something to the application that is using it. Usually these are running to VMs or containers and deployed to the environment where they can be accessed, ie. cluster. Baas (Backend-as-a-service) is a more broad concept of functionality that applications can use. Baas can include databases, pub-sub queues and many other higher level functionalities. The main point is that developers don't need to set these up themselves but can just use the platform providers own solution. These relate to my definition neatly, because developers don't need to know how they work, but just need to know how much it costs to use these solutions.

4

"Scale to zero" means that when no one is using your application there is no need and sense to keep it running in the server. This also means that developers are not charged of the usage of the server when it is not running on the server. When someone requests the service via HTTP -request for example, then it is started again in the server.

5

"Cold start" means the time when user requests serverless service but which isn't running on the server yet. Server platform needs to start the service and possibly handle all kinds of stateless initializations of other services depending on if the service uses any of these. On the meantime user just needs to wait until the service is running. This could take several seconds.

6

Figure 1 shows the way new videos are cut into segments and segments are stored to other locations. When user wants to watch the videos these segments are then transcoded into suitable format for the user to watch.

Reason netflix wants to transcode the video files is that there are multiple video players that support different video formats. By transcoding the video more people can watch these videos.

Transcoding can also reduce the file size of the videos, which in return allows more traffic to go through netflix's servers allowing more people to access the service.

7

i)

OVF is a example of a cloud computation standard. It's aim is to standardize the creation and management of virtual machines in the cloud.

https://www.dmtf.org/sites/default/files/DMTF_OVF_2.0_FAQ_Final_1.pdf

<https://www.dmtf.org/standards/ovf>

Other than standards there are also many rules/recommendations of independent cloud providers to tell the best way to use their services. AWS lambda rules <https://awslabs.github.io/serverless-rules/rules/>

ii)

Companies might not want to standardize their technologies. New technology invented inside a big tech/telecom company is always a leverage to gain upperhand on competition. Companies might not want to share these things to competitors.

8

It is related to McCarthy's vision because McCarthy suggested that in the future subscribers will only pay from the capacity they use but has access to all the utilities available. This ties neatly to the idea of serverless that if no one uses your app it is not running on the server and you will not be charged for it.

9

Reducing cold starts is needed because with minimal cold starts serverless could become much more performant. This ties to the McCarthy's idea of public utility because to be considered public utility means the usage is much more higher than it is now. If user have to wait for services to start usage of the service would decrease.

Cloud computing could possibly reduce cold starts by moving idle services from the cloud to edge devices where they won't stress the server so much. When traffic to that service increases then it is moved back to server where it can handle increased traffic.

Service-level agreement is needed. Without the quality of service there will not be customer trust to build other technologies atop of serverless.

10

Serverless cannot be considered yet as public utility because there are no regulations or laws set to serverless providers which would guarantee the availability of the service. As it is now serverless services are just commercial products provided by certain companies and if they deem so they could stop providing the service. To make it public utility there would be need to regulate it so that that cannot happen, to ensure the availability.