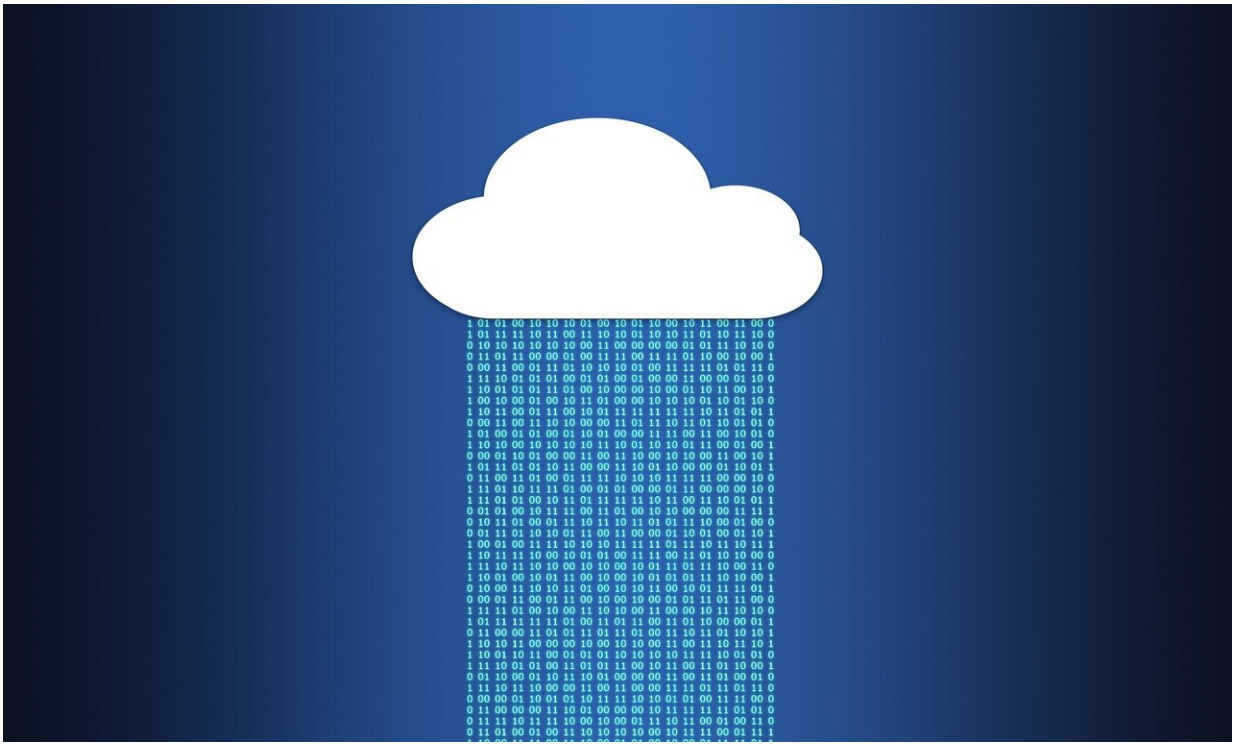
Cloud services without servers: What's behind it

September 20 2023, by Robert Emmerich



Credit: CC0 Public Domain

A new generation of cloud services is on the rise. It is based on the paradigm of "serverless computing," which is an active research topic at the Institute for Computer Science in Würzburg.

In cloud computing, commercial providers make computing resources

available on demand to their customers over the Internet. This service is partly offered "serverless," that is, without servers. How can that work? Computing resources without a server, isn't that like a restaurant without a kitchen?

"The term is misleading," says computer science Professor Samuel Kounev from Julius-Maximilians-Universität (JMU) Würzburg in Bavaria, Germany. Because even serverless cloud services don't get by without servers.

In classical cloud computing, for example, a web shop rents computing resources from a cloud provider in the form of virtual machines (VMs). However, the shop itself remains responsible for the management of "its" servers, that is, the VMs. It has to take care of security aspects as well as the avoidance of overload situations or the recovery from system failures.

The situation is different with serverless computing. Here, the cloud provider takes over responsibility for the complete server management. The cloud users can no longer even access the server, it remains hidden from them—hence the term "serverless."

"The basic idea of serverless computing has been around since the beginning of cloud computing. However, it has not become widely accepted," explains Samuel Kounev, who heads the JMU Chair of Computer Science II (Software Engineering). But a shift can currently be observed in the industry and in science, the focus is increasingly moving towards serverless computing.

A recent article in *Communications of the ACM* deals with the history, status and potential of serverless computing. Among the authors are Samuel Kounev and Dr. Nikolas Herbst, who heads the JMU research group "Data Analytics Clouds."

Experts define serverless computing inconsistently

The origins of the research article lie two years back. In 2021, around 50 international experts in serverless computing met at a seminar at Schloss Dagstuhl—Leibniz Center for Informatics.

"We discussed the most important developments and research questions there and found out that many of us define serverless computing differently and that contradictions sometimes arise," Kounev recounts. This is not unusual in science, he says, when a paradigm shift emerges.

In order to create clarity, a small group of researchers got together. Together, they tried to define serverless computing precisely after the seminar. Exactly this circle of people has now published the article in the renowned CACM magazine; it also includes Ian Foster, a prominent pioneer of cloud computing from the U.S.

Two key principles defined

The team has defined two principles that characterize serverless computing. "NoOps" is the first, which stands for "no operations." This means, as described above, that the technical server management, including the hardware and software layers, is completely in the responsibility of the cloud provider.

The second principle is "utilization-based billing," which means that only the time during which the customer actively uses the allocated computing resources is billed. With classic cloud services, on the other hand, a fee is paid for the entire time during which the respective cloud resources were rented.

At JMU, the computer scientists in Samuel Kounev's team are working, for example, on the elasticity of cloud services, especially the

autoscaling of computing resources. The goal here is that the amount of computing resources allocated to a cloud application automatically adapt to increasing or decreasing demand over time.

A long-term project that Kounev's team wants to realize in the coming years is a serverless cloud platform for large workflows in Earth observation. The focus is on climate research with satellite data. Scientists should be able to use the platform to study various global effects of climate change easily, quickly and efficiently. These include changes in forests, snow cover in mountains, or biodiversity.

In this field, Kounev's team is cooperating with the JMU Chair of Remote Sensing, the German Aerospace Center (DLR), the Leibniz Computing Center of the Bavarian Academy of Sciences and Humanities and the Max Planck Institute for Behavioral Biology (Radolfzell / Konstanz), among others.

More information: Samuel Kounev et al, Serverless Computing: What It Is, and What It Is Not?, *Communications of the ACM* (2023). DOI: 10.1145/3587249

Provided by Julius-Maximilians-Universität Würzburg