

Lesson 3

Reading: "History of the cloud"



Blesson Varghese, Assistant Professor at Queen's University Belfast, traces the footsteps of cloud technology, starting from its foundations five decades ago, through its development, to examine just where the cloud is heading in the future.

The current digital age is marked by our reliance on the cloud. Every internet user is confronted with a variety of cloud choices and options, which was simply not available a decade ago. A plethora of services we rely on either for downloading apps or for storing our photos and videos are hosted by the cloud. The frustrations we have, when we cannot connect to the cloud, are evidence of how much it has become integral to our lives. But, how did we get here? What is the history of the cloud we need and now use every day?

The foundation era - 1960s

A mashup of three fundamental concepts define the cloud: the first is delivering a service, such as computing or storage as a utility; the second is multiple people sharing the same computer resource, which is possible through a technology, referred to as virtualisation; the third is accessing services via networking. The 60s was a momentous decade in laying this foundation.

The late Professor John McCarthy - a visionary computer scientist who coined the term 'artificial intelligence' - suggested in 1961 that computing would one day be sold as a utility. In 1967, IBM virtualized operating systems allowing for multiple users to timeshare the same resource. In 1969, ARPANET (Advanced Research Projects Agency Network), a network relying on the TCP/IP protocol was launched by the US Department of Defense and became the forerunner of the Internet.

The melting pot - 1970s/80s

Research in virtualisation, operating systems, storage and networking advanced in the next two decades finding new applications never known before. This melting pot fused technologies for novel outcomes, such as networks that allowed wire transfers between financial institutions. In the UK, scientists worked on using unused television signals for sending data.


In 1976, advances in networking was demonstrated by Queen Elizabeth II by sending an email. A couple of years later, the White House installed its first computers. In the early 80s, network operating systems were launched to allow computers to talk to each other. By 1985, storage tapes that could store up to 200 megabytes of data (an average smartphone has 10 times this memory) were available. By this time around 100,000 computers were connected to the Internet.

The march begins - 1990s


The foundational technologies for the cloud reached a certain level of maturity in the 90s. This was epitomized by the launch of the World Wide Web in 1991 when more than a million machines were connected to the internet. This led to the dotcom revolution and e-commerce gaining popularity. The client-server model of distributed computing was implemented, such that websites provided front-ends for users and servers located in the World Wide Web hosted the backend logic.

The earliest mention of cloud computing in literature is known to appear in an internal document of Compaq in 1996. In 1997, Professor Ramnath Chellappa from Emory University had mentioned the cloud in an article. Although the term 'cloud' was coined, it was a predecessor of cloud computing - known as grid computing - that became popular. Although grids networked compute resources of organizations across the continents, it was still not accessible to non-specialist users or developers.

The inception - 2001-04



The inception of the modern-day cloud was realized by Amazon Web Services (AWS) launching its public cloud in 2002. There were virtually no competitors at this stage and while the benefits of using the cloud, such as elasticity and scalability were known, the real use cases to convince potential users were not yet available.




Indeed, the cloud offered a dream solution to the technical and management nightmares that many small and medium scale businesses and organizations suffered. The cloud alleviated the burden of maintaining servers, upfront investment on compute resources and scaling web services based on demand - by renting resources that were concentrated in large facilities and maintained by trusted providers.

While more websites and workflows were starting to find a place in the cloud, the next decade saw the cloud develop over two distinct generations.

The first generation cloud - 2005-11

The traditional definition of what the cloud is was realized during the first generation - centralized infrastructure in data centers that host a lot of compute and storage resources. The momentum gained during this time allowed application owners to typically exploit a traditional two-tiered architecture in which cloud providers hosted the backend, whereas the users sent all their requests from web and mobile applications, to the cloud.



The OpenNebula research project was launched to develop an entire software stack for easy use of the cloud in 2005. The Elastic Compute Cloud (EC2) from AWS was made available to the general public in 2006. During this time, the technology was further developed and solutions focused on empowering data centers - it made applications reliant on them while mitigating risks for the user.

Database services started becoming available on the cloud and the popular Dropbox concept facilitated cloud storage as a service. Cloud providers offered data centres in a few locations, but after 2008, multiple regions were available. Service level agreements (SLAs) were articulated to quantify and guarantee a quality-of-service (QoS) to the user.

- As the software stack required institutions to organize their clusters as clouds became available, private clouds started to emerge. It was not until 2009 that a standards body provided a definition and later still in 2011 when a reference architecture became available. Microsoft entered the marketplace, launching Azure services. Similar to OpenNebula, the OpenStack software was launched in 2010. Both these projects were open source and attracted huge interest from the IT community.

The second generation cloud - 2012-17

Building on the first generation cloud, the second saw a dramatic enrichment in both the services provided and increased competition with a wide choice of providers. As it was possible to monitor the resources used on the cloud, the concept of the cloud became more trusted. In addition to the pay-as-you-go pricing model, spot bidding was introduced for resources. Real-time streaming services started processing data on the cloud.

In addition to relational databases, non-relational database services were now available. DevOps gained prominence and microservices were used for cloud application development. This was fuelled by the launch of container services on the cloud in 2014. A newer definition started to emerge where public and private clouds could be combined to deliver hybrid clouds. In Europe, the European Grid infrastructure federation cloud was launched in 2014.

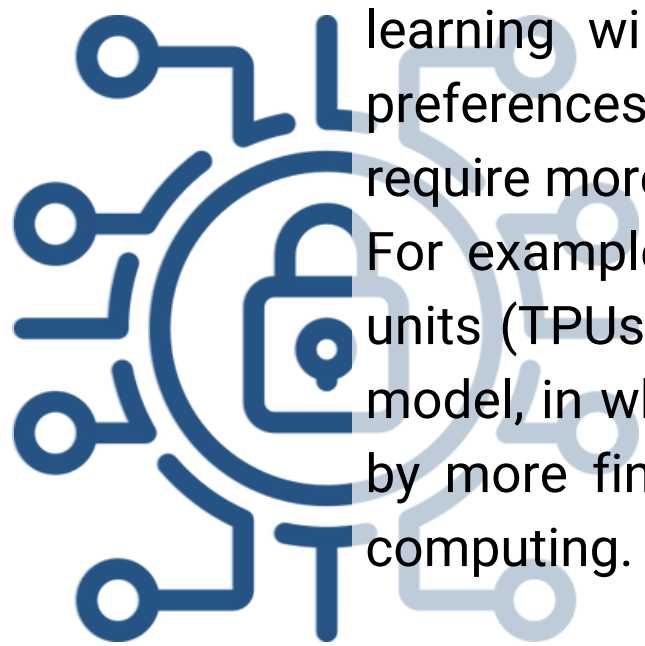
What's going to happen next? (2017 and beyond)

Two significant developments during the first and second generation cloud implementation initially went unnoticed. The first was related to processing outside the cloud and the second regarding heterogeneous cloud resources.

While centralized clouds (which were distant from end users) had gained popularity, research had begun on whether some of a user's requests could be processed outside the cloud to reduce communication latencies. The 'cloud in a box' concept emerged. The possibility of computing on networking elements, such as routers and base stations was defined as 'edge computing' in 2009. Later on, edge computing received further attention from Cisco under the umbrella of 'fog computing,' in which computing is facilitated along the entire cloud-edge continuum.

The underlying hardware in data centers became more heterogeneous. Usually, applications running on the cloud use processors, such as CPUs, which are seen in desktop PCs. However, many of them need to execute faster, so accelerators such as graphics processing units (GPUs) were added to data centers.

With the advent of the internet-of-things era, billions of devices are anticipated to be connected to the internet. They cannot all send data to the cloud due to increased network traffic and communication latencies. Fog/edge computing is a good mechanism to alleviate these challenges and therefore the market seems to be embracing the edge technology. Micro data centers are now launched undersea.



Methods for making the devices more secure when computing will be done on the network edge, such as using blockchains will gain prominence. More machine learning will be required in the cloud to predict user preferences and understand diverse workloads. These will require more specialist processors suited to the workload. For example, Google has incorporated tensor processor units (TPUs) in the cloud. The coarse, pay-as-you-go cost model, in which a user pays for idle time, will be replaced by more fine-grain billing models, such as in serverless computing.

Although the cloud is currently crowded, the technology offers more room for expansion. The golden era of cloud computing is yet to dawn on us.

Taken from: <https://www.bcs.org/articles-opinion-and-research/history-of-the-cloud>

Fill in the blank activity: "History of the cloud"

Para consolidar las respuestas ingrese al cuestionario online.

Kahoot activity based on the vocabulary and comprehension of the text.

LINK: <https://play.kahoot.it/v2/?quizId=bbc0dc44-cda7-430c-badd-5af37625f115>