

Mei Hong

# Cloud Computing: Ten Years and Beyond

| By *Mei Hong, Member of the Chinese Academy of Sciences and Professor of Computer Science and Technology, Peking University*

*This review is based on Professor Mei Hong's keynote speech at the 2016 China Cloud Computing Conference in Beijing in which he examined cloud computing technologies to date and invites us to take a look into the future. >>*



The year 2006 is regarded as ‘year one’ for cloud computing. Using virtualization technology, Amazon pioneered the cloud-based Hardware-as-a-Service (HaaS) model, which was designed to deliver computing resources to the public similarly to the way utilities distribute water and electricity.

Since then, cloud computing has become the primary infrastructure for Internet innovations and ubiquitous computing — converging society, information, and the real world. Along the way, this trend has demanded new computing models and platforms. To put this in perspective, we will look at the three phases of cloud development.

## Phase 1 (2006 to 2010): The Basics

As companies competed to offer public cloud products, the HaaS model was widely recognized but not necessarily well understood. Unfortunately, HaaS did not come with a clear definition of cloud computing. Industry giants and researchers provided their own explanations, but all fell short. Hindsight reveals that the virtualization of large-scale computing resources and service-oriented software stacks were the key

enabling technologies for what cloud computing has become.

Hardware resource virtualization and management technologies boomed and deepened our collective understanding of the cloud. Other major technical developments for cloud computing have included the following innovations:

- 2007: Kernel-based Virtual Machine (KVM) merged with Linux kernel mainline
- 2008: Linux Containers (LXC) released
- 2009: VMware vSphere launched
- 2010: CloudStack open-source released

In 2010, NASA and Rackspace jointly launched the OpenStack project, which has become important for the development of private cloud infrastructures. Innovative service models have emerged and, in turn, have bred numerous concepts featuring X-as-

a-Service (XaaS). The release of important cloud computing technologies, such as open-source software, has become a defining characteristic of the cloud-computing infrastructure.

## Phase 2 (2010 to 2015): Getting on Board

Cloud service providers captured market share and fought for competitive advantage. Rapid adoption produced a global market worth about USD 100 billion. In September 2011, the United States National Institute of Standards and Technology (NIST) released a white paper, which provided an authoritative definition of cloud computing that has since been widely accepted. XaaS was primarily available in three forms:

- Infrastructure-as-a-Service (IaaS)
- Software-as-a-Service (SaaS)
- Platform-as-a-Service (PaaS)

Despite starting later, the market for private clouds developed at a faster pace than the public market. Private-public hybrid clouds also emerged, which served to increase mobility and terminal connectivity to further shape the deployment of cloud computing services.

As key technologies and systems for cloud services and management matured, open-source computing platforms, such as OpenStack and CloudStack, gained wide deployment. Software-Defined Networking (SDN), represented by OpenFlow, also became an important enabler. Industries reached a consensus on the application of software-defined hardware resources, including computing, storage, and networking, to meet the growing demand for deploying large numbers of Virtual Machines (VMs) with increased flexibility. Software-defined management platforms began to allow for efficient control over the now massive scale of cloud computing resources.

## Phase 3.0 (2015 to 2020): In Full Swing

Today, the maturity of cloud platforms and the emergence of a suitably large number of terminals have created a ‘prosperous development’ phase for cloud computing. Importantly, cloud-computing service providers have shifted their focus from cloud

facilities to applications. Providers’ current concerns are to meet the growing diversity of user-application requirements that have given rise to a marketplace for Application Programming Interfaces (APIs).

Big Data is an important feature of cloud computing, with spending on cloud-based Big Data and Analytics (BDA) expected to grow faster than that for on-premises solutions. An increasing number of cloud-based, device-oriented applications are expected to meet the requirements of professional users. Based on software-defined platforms that deliver greater flexibility and support, cloud-device convergence will become the new model for cloud computing.

## Envisioning Future Development

The future of cloud computing will center on ubiquitous resources, platform-based systems, industry-specific applications, and improvements in service quality. As more devices are cross-connected through the Internet, society, information systems, and the physical world will increasingly converge.

Ubiquitous resources refer to various computing resources for further expansion, such as mobile Internet clouds, smart terminals, and nodes on the Internet of Things (IoT). New cloud computing architectures that feature the convergence of the cloud with devices are being developed for the on-demand allocation of client and server resources. Future clouds must not only support existing terminals, such as smartphones and tablets, but also various IoT-connected devices. A serious challenge for ubiquitous connectivity will be to manage huge amounts of cloud-device resources.

Timely utilization of massive sensor architectures is vital for cloud platform management. Public, private, and hybrid clouds of varied scales and build characteristics will emerge:

- Large-scale clouds will provide services to users worldwide.
- Small clouds will be built on existing resources.
- Physical clouds will possess their own hardware resources.
- Federated cloud alliances will be built on a foundation of physical clouds.



*Today, the maturity of cloud platforms and the emergence of a suitably large number of terminals have created a ‘prosperous development’ phase for cloud computing. Service providers have shifted their focus from cloud facilities to applications. >>*



**The future of cloud computing will center on ubiquitous resources, platform-based systems, industry-specific applications, and improvements in service quality. >>**

Demands for cross-cloud alliances are growing. For cloud service providers, serving applications across multiple clouds and realizing open collaboration and in-depth cooperation between cloud providers are important issues that must be addressed.

Platform-based systems are a sign that cloud computing support systems are evolving to cloud Operating Systems (OSs). Although much discussed, cloud OSs have yet to attain the level of architectural maturity and capability that are expected. Current systems are primarily responsible for managing cloud resources to support various applications. Future systems will also need to manage multiple tasks running on clouds, similar to standalone OSs. An ideal cloud management system should include cloud OSs, stand-alone OSs, and various application containers and middleware to support a wide variety of cloud services.

To begin, cloud OSs must seamlessly port traditional applications in order to manage massive numbers of heterogeneous resources effectively. Future OSs must support integrated interactions between dissimilar terminals connected through the Internet.

Furthermore, cloud OSs must better support uplinks to applications. Developers are encouraged to explore the construction techniques and operational impact of cloud-native applications, research and develop new programming design models and related languages, design unified scheduling and management mechanisms for cloud tasks, and implement on-demand integration of resources within and across clouds.

A critical enabler of cloud resource management, Software-Defined X (SDX), supports custom development based on user requirements. The promise of SDX is extensive management support with the ability to cover micro, mid-range, and large-scale VMs.

Industry-specific applications are geared toward the provisioning of API-integrated environments by cloud service providers to address specific solutions and connectivity domains.

Software-service awareness is an important

enabler for industry-specific cloud applications. Early information systems were tightly coupled and integrated, comprising applications developed by one vendor that did not support either custom development or interoperation with third-party applications. Software-Oriented Architecture (SOA) from the 1990s gave rise to SaaS which, in turn, is able to support loosely coupled distributed applications based on coarse-grained Internet services.

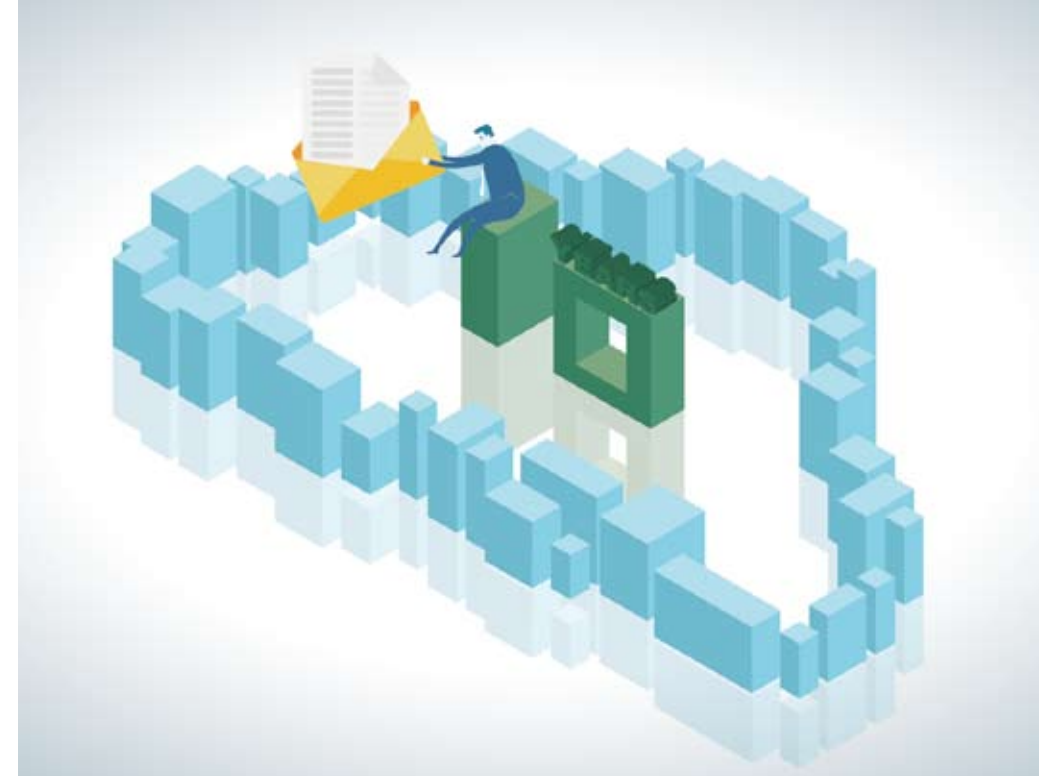
Improvements in service quality can be described in three words: higher, faster, and stronger.

Higher refers to greater throughput and the aggregation of massive numbers of support data processing with excessively high concurrent access — a common requirement of many cloud applications.

Future clouds must respond faster to improve the user experience and service quality. Amazon has found that every 100 ms of latency costs 1 percent in sales. A one-second delay in page response can significantly reduce customer conversion rates, page visits, and customer satisfaction. By comparison, Augmented Reality (AR) and Virtual Reality (VR) require 1 ms response times to satisfy customer requirements.

The key to strong, high-quality applications in the cloud is fast response to service requests. Cloud-based computing applications are subject to latencies from two main sources: the network and the cloud data center. Current statistics reveal that the network and data center each contributes about half the latency that users experience during service usage. Network latency can be reduced by higher bandwidths and better geographic distribution of data centers for user proximity. Latency caused within data centers can be reduced through better vertical integration of layered cloud-aware software stacks.

As the requirements for cloud services become more diversified and attract more organizations to switch to cloud computing platforms, researchers will continue working on initiatives in areas like VM status synchronization, parallel data and graphics calculation, large-scale nonvolatile



memory, and distributed uninterruptible power supply systems.

### 'Internetware' for Cloud Computing

Researchers in China have proposed the term 'Internetware' to describe Internet-related software with new features that would change current software models, operating platforms, engineering, and quality.

Sponsored by *National Basic Research Program 973*, the researchers have built an open, collaborative Internetware model. The first model has been operating at Peking University for several years. There, a software team has pioneered research in hybrid cloud management, data interoperability platforms, and cloud-enabled Big Data processing. Based on the guiding principle of 'Whole Software Architecture Lifecycle,' they have further proposed a container structure and associated mechanisms to support on-demand collaboration and online evolution that will enable autonomous system management and operational support.

In the context of hybrid cloud management R&D, the team has adopted a software-defined approach that features API-based management functions and programmable management tasks in a system called YanCloud IaaS. This system supports integration and configuration management of infrastructure resources and the on-demand construction and management of public, private, and hybrid IaaS clouds. A number of Chinese IT companies have developed their own cloud management products based on YanCloud IaaS for eGovernment, transportation, telecommunications, and health. In 2015, the Internetware research program won a *Science and*

*Technology Progress Award* from the Ministry of Education.

The software team developed the YanCloud Data-as-a-Service (DaaS) system to resolve interoperability issues that hindered data sharing over the Internet. YanCloud adopts a structure recovery technology that captures APIs and their system-facing components to enable interoperability between applications and data to form new operational management logic. The YanCloud DaaS system encapsulates data from various web systems, mobile applications, and PC software into APIs without needing the original documents or system source code. The system has enabled data sharing for over a hundred government, finance, transport, and energy application environments.

For cloud-enabled Big Data processing, the software R&D team has developed a lightweight data management and processing platform called Docklet. Docklet is a cloud OS for mini data centers where each user has a private Virtual Cluster (VCluster) of Linux container nodes mapped onto distributed physical machines. Each VCluster is separate and can be operated like a physical cluster. Docklet supports various computing frameworks, including Spark and MPI, and can run data analytics and processing programs in Python, R, and Java. Docklet offers users cloud-based workspaces with many programming frameworks preinstalled. Browsers are used to complete all data analytics operations, including editing, debugging, and programming. At the Computer Center of Peking University, Docklet provides teachers and students with a variety of cloud services for scientific computing, data analysis, visualization, and virtual experimental environments. ▲



**Researchers in China have proposed the term 'Internetware' to describe Internet-related software with new features that would change current software models, operating platforms, engineering, and quality. >>**