Huawei Technologies Co., Ltd., *Cloud Computing Technology*
https://doi.org/10.1007/978-981-19-3026-3_1

1. Introduction to Cloud Computing Computing

Huawei Technologies Co., Ltd.[1] ✉
(1) Hangzhou, Zhejiang, China

This chapter is an overview of cloud computing, including common cloud computing scenarios in life, the characteristics, definitions, origins and development of cloud computing, the advantages and classification of cloud computing, various supporting technologies of cloud computing. The business model, computing model, and implementation of cloud computing are three perspectives for understanding, as well as open source methods that are currently very popular in the cloud computing field. Through the study of this chapter, I hope that readers have a clearer understanding of the general picture of cloud computing and lay the foundation for the in-depth study of the following chapters.

1.1 Ubiquitous Cloud Computing

As a representative of a new technology, cloud competing, like internet, has closely penetrated into our daily live. For example, we want to share an electronic material of hundreds of Mb with a friend from distance place, what happens if it exceeds the limitation of email attachment size? In the past, we generally used express delivery of storage media such as CDs, flash drives, or mobile hard drives which is time consuming and cost more work. Now we have a much more convenient way with the help of cloud storage service such as Baidu disk. Just put the data file into your own cloud disk and send the sharing link and access password to the recipient. The recipient can obtain the shared data file anytime and anywhere via the Internet. Another example is that an organizer wants to hold a special meeting while the participants are located all over the country. In an epidemic prevention and control situation, having participants gather by transport from all over the country for an on-site meeting not only takes considerable time and expense to travel back and forth, but also increases the risk of spreading the epidemic. Therefore, people will prioritize ZOOM meeting, Tecent meeting, or Webex as an option to hold online meeting. Participants only need to use the Internet to perform simple operations using a browser, and they can quickly and efficiently share voice, data files, and videos with participants in different geographical locations. In fact, participants in a cloud conference only need to have a device (computer, mobile phone, tablet, etc.) that can access the Internet that can be used normally to achieve

online communication and video conferences without having to care about the complex technologies such as data transmission and data processing, all of which are provided by cloud conference service providers.

Such a way of preparing resources in advance and using these resources to perform specific tasks through specific technologies anytime and anywhere is generally a cloud computing type. The provider is a cloud service provider such as Huawei's public cloud. Let's take a look at the Huawei cloud website as shown in Fig. 1.1.

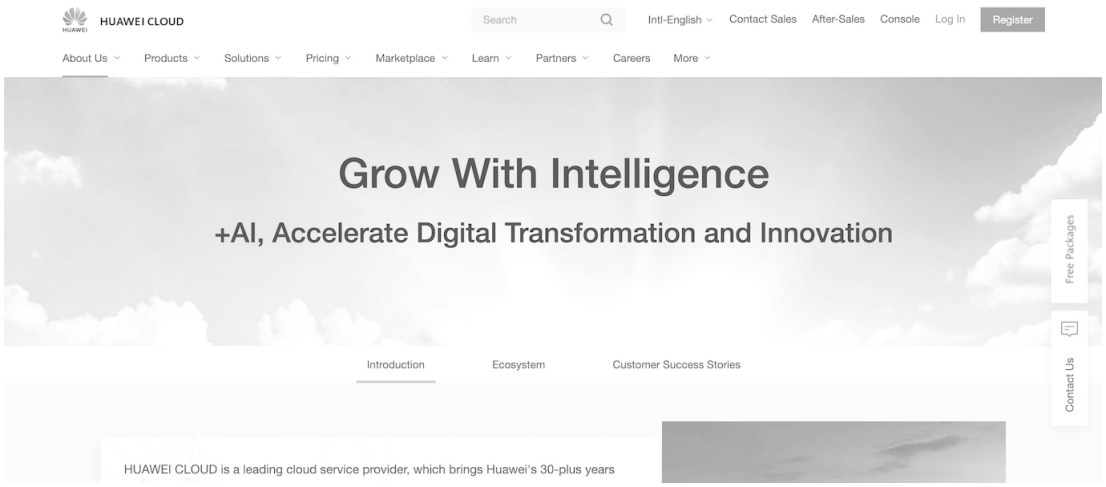

Fig. 1.1 Huawei cloud website

Under "Product" → "Fundamental services", we can find computing, storage, network, database, container services, etc. These divisions can be divided into different subdivision types. Take a popular service—ECS, an elastic cloud server, as an example, as shown in Figs. 1.2 and 1.3.

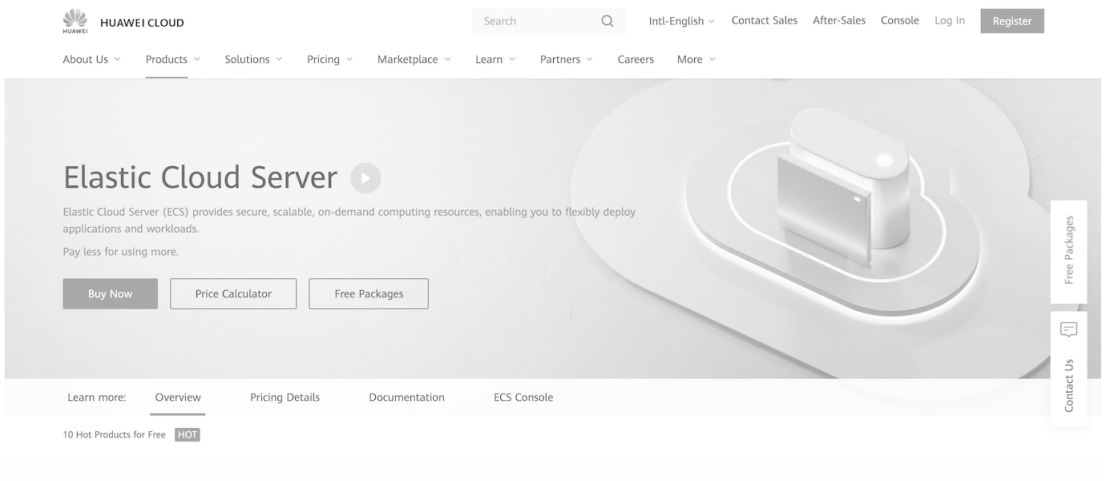

Fig. 1.2 Elastic cloud server

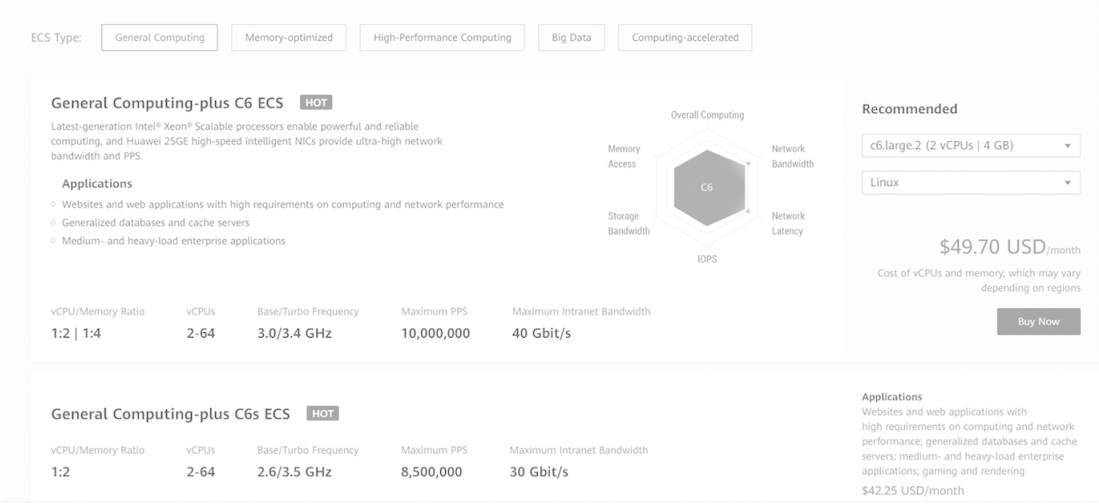

Fig. 1.3 Elastic cloud server specifications

The elastic cloud server on the website is actually a virtual server (we will introduce it later). Similar to our own purchase of computer, the website provides different grades and types of cloud server instances to choose from. The configuration includes parameters such as the number of CPU, frequency, memory, and network bandwidth. Users can choose the most effective cloud server according to their needs. In fact, buying a cloud server instance is like buying a physical machine. You can complete most of the work that can be done on a physical machine, such as editing documents, sending emails, or working together. It's just that the cloud server is not in front of you, but on the far end of the network (cloud). In addition, the cloud server also has some advantages that the local physical machine does not have. For example, the access to the cloud server is not restricted by time and place. As long as there is Internet, it can be used anytime and anywhere. And the equipment (terminals) for operating the cloud server can be varied. For example, the user can operate the cloud server through a personal computer (PC), mobile phone, etc., and can modify or expand the performance configuration of the cloud server if necessary.

In addition to providing cloud servers to users, cloud service providers generally provide some other cloud services. For example, on Huawei's public cloud, users who need to build a website can purchase the cloud speed website building service, which can help users quickly complete the construction of the website; users who need to store data can purchase object storage services or cloud hard drives. More advanced services also include artificial intelligence (Artificial Intelligence, AI) functions such as face recognition, voice recognition, image recognition, or text recognition.

In short, cloud computing allows us to use IT services like water and electricity as soon as the user turns on the faucet, water rushes out. This is because the water plant has sent water into the pipeline network (water network) that connects thousands of households; electricity is similar. For cloud computing, cloud service providers have prepared all resources and services for users, and users can use

them via the Internet. The Internet here is similar to the previous water network, and the tap can be a browser or a mobile application (App).

In fact, cloud services around you can be seen everywhere. In addition to the examples introduced in the previous article, other cloud services such as automatic backup of mobile phones, Youdao Cloud Notes and NetEase Cloud Music are all cloud services around us. At present, mainstream mobile phone manufacturers such as Huawei and Apple provide cloud-based mobile phone backup and recovery services. Users can back up files on the mobile phone to a data center in the cloud. After replacing the phone, you can restore your data to the new phone using your account and password. Youdao Cloud Notes is a product launched by NetEase, which provides online document creation and editing functions. When the user needs to record his inspiration at a certain moment, but unfortunately finds that there is no paper and pen around, the user can use Youdao Cloud Notes to record the inspiration online. Another advantage of this product is that no matter when and where, no matter what terminal the user uses (personal computer, mobile phone, etc.), online data can be edited, shared, and collaborated anytime, anywhere, and every edit can be done immediately Sync to the cloud. Music lovers may like NetEase Cloud Music App, through which songs can be listened to online and played at any time.

With the rapid development of cloud computing technology, similar cloud services will increasingly penetrate our daily lives. The spring weather turns rain, moisturizing things silently. We can truly feel the convenience of cloud computing technology in our lives.

1.2 The Properties of Cloud Computing

Cloud computing, as a new computing model, mainly has the following characteristics.

1.2.1 On-Demand Self-Service

Speaking of on-demand self-service, the first thing comes to your mind is supermarket. Every customer can collect goods according to their requirements. If it is the same type of goods, you can check description, price, and brand information to decide whether to buy or which one to buy. On-demand self-service is one of the mail characteristics of cloud computing. We will later introduce the Infrastructure as a Service (IaaS), platform as a service (PaaS) and Software as a Service (SaaS) model. Users can choose among one of these models based on their necessity. After selecting the mode, there will generally be different configurations to choose from, and users can purchase the services according to their needs. The entire process is generally self-service and does not require third-party intervention unless you have a problem that requires consultation. As shown in Fig. 1.3, Huawei’s elastic cloud server specifications for public clouds have many different configurations of cloud server instances to choose from.

On-demand self-service is premised on knowing your needs and which products will address them. This requires the relevant expertise of users using cloud computing. Users who do not have the knowledge and capabilities to use cloud services can consult a cloud service provider or turn to a relevant professional services provider.

1.2.2 Extensive Network Access

Another feature of cloud computing is that all clouds must rely on network connectivity. It can be said that the network is the foundation of cloud computing. Especially the Internet, the cloud is always inseparable from the Internet. The Internet provides remote, anytime, anywhere access to IT resources. Some people even think of cloud computing as “Internet plus computing,” and network access is an intrinsic property of cloud computing.

Although most cloud access is over the Internet, cloud users also have the option of using a private channel to access the cloud. The level of service for network connectivity between cloud users and cloud service providers (quality of service, Quality of Service, QoS) depends on the Internet Service Provider (ISP) that provides them with network access.

In today’s society, the Internet can cover almost every corner of the world, we can connect to the Internet through a variety of digital terminals, such as personal computers and mobile phones and connect to the cloud through the Internet, using cloud services. Therefore, extensive network access is an important feature of cloud computing. This can either be a wired network or a wireless network such as a Wi-Fi network. In short, without the network, there would be no cloud computing.

1.2.3 Resource Pooling

Resource pooling is one of the prerequisites for on-demand self-service, through resource pooling can not only put similar goods together, but also can refine the units of goods. Slightly large-scale supermarkets will generally be divided into fresh areas, fruit and vegetable areas, daily necessities areas and other areas to facilitate customers to quickly find their own needs of goods, but this form is not a pool of resources, can only be regarded as a classification of resources. So what is pooling resources? In addition to converting similar resources into resource pools, resource pooling requires the decomposition of all resources into smaller units. If we buy our own hard drives, a mechanical drive (Hard Disk, HDD) often has a few terabytes (TB, 1TB, 1012B); solid-state drives (Solid State Drive, SSDs) have a slightly smaller capacity, and an SSD typically has a capacity of 128 to 512GB (Gigabytes, 1GB, 109B). Storage pooling cannot be measured in the number of hard drives because a hard drive has a large capacity, some applications only need a few gigabytes (GB), allocating the capacity of a hard disk is obviously a huge waste. Therefore, the way to use resource pooling need to break the number of physical hard disk unit

"one" and combined all the capacity of the hard disk, gathered into a "pool." Then allocation can be assigned in smaller units such as "GB" as a unit. Users can apply for as much as they need.

The computing resources include CPU and memory. If the CPU is pooled, the smallest unit of the CPU that the user sees can be a virtual core, and the CPU manufacturer no longer reflects the physical attributes of AMD or Intel.

Another function of resource pooling is to screen the differences between different resources. After the storage resources containing the mechanical hard drive and the SSD are pooled, if the user requests a certain amount of storage space, which corresponds to the mechanical hard drive or SSD, or both, he cannot tell the difference. In cloud computing, resources that can be pooled include computing, storage, and networking. Computing resources include CPU and memory. If CPU is pooled, the smallest unit of the CPU that the user sees can be a virtual core, and no longer reflect physical attributes such as the CPU's manufacturer being AMD or Intel.

1.2.4 Fast and Elastic Scaling

Fast elastic scaling is one of the characteristics of cloud computing and is often cited as one of the core reasons for attracting users to "embrace" cloud computing. Cloud users can automatically and transparently scale their IT resources according to their needs. For example, in order to deal with the sudden high traffic of hot events, users can temporary self-purchase a large number of virtual resources to expand capacity. When hotspot events "cool down" and access traffic tends decline, users can release these newly added virtual resources, which is typical of fast elastic scaling. Cloud providers with large IT resources can provide a wide range of elastic scaling.

Fast elastic scaling includes several types, and in addition to manual capacity expansion or reduction, cloud computing supports automatic scaling or reduction based on preset policies. Scaling can be an increase or decrease in the number of servers, or an increase or decrease in resources for a single server.

In cloud computing, the biggest benefit of fast elastic scaling for users is cost savings while keeping the business or application running smoothly. Enterprises can purchase small amounts of resources when they are in low initial demand, gradually increase their investment in resources as the size of the enterprise expands, or concentrate all resources on priority business use during special periods, and, if resources are not sufficient, immediately apply for additional resources and, after a special period, release new resources. Either scenario is convenient for the user.

1.2.5 Measurable Services

Measuring is not billing although measuring is the basis of billing. Among the services provided by cloud computing, most services need to be paid for, but there are also services that are free. For example, elastic scaling can be opened as a free service for users.

Metrology is the use of technology and other means to achieve unity and accurate and reliable measurement. It can be said that the services in cloud computing are all measurable, some are based on time, some are based on resource quotas, and some are based on traffic. Measuring service can help users to automatically control and optimize resource allocation accurately according to their own business. In cloud computing systems, there is generally a billing management system that is specifically used to collect and process usage data. It involves the settlement of cloud service providers and the billing of cloud users. The billing management system allows for the development of different pricing rules and can also customize the pricing model for each cloud user or each IT resource.

Billing can choose between prepaid use or pay after use. The latter payment type is divided into predefined limits and unlimited use. If the limit is set, they often appear in the form of quota. When the quota is exceeded, the billing management system can reject the cloud user's further use request. Assuming that a user's memory quota is 500GB, once the user's storage capacity in the cloud computing system reaches 500GB, new storage requests will be rejected.

Users can purchase services according to their needs and can clearly see the usage of their purchased services. For contract users, the type of product used, service quality requirements, cost per unit time, or cost per service request are usually specified in the contract.

Figure 1.4 shows the pricing standards of Huawei Elastic Cloud Server instances, which shows the pricing standards of virtual server instances with different configurations. In this example, they are charged monthly.

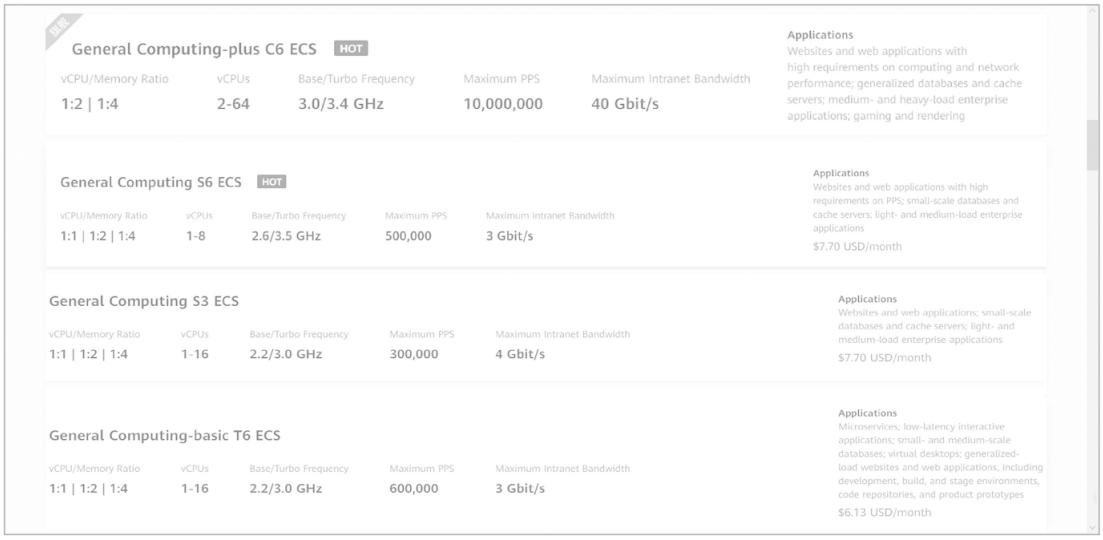

Fig. 1.4 Pricing standards for Huawei Elastic Cloud Server Instances

1.3 Definition of Cloud Computing

There are several definitions of cloud computing. There are many definitions of what cloud computing is.

Wikipedia: Cloud computing is an Internet-based computing method. In this way, shared hardware and software resources and information can be provided to computers and other devices on demand, just like water and electricity for everyday use, paid for on demand, without caring about their source. National Institute of Standards and Technology, NIST: Cloud computing is a pay-per-use model that provides usable, convenient, on-demand network access to configurable computing resource sharing pools (resources including storage, software, services) that can be delivered quickly with minimal administrative effort or little interaction with service providers.

In the past, engineers used to use clouds to abstractly describe telecommunications networks or the Internet and underlying infrastructure when drawing pictures. The name of cloud computing has an inextricable origin. The “cloud” in cloud computing can be seen as a vast pool of IT resources where users can purchase the services they need on demand and pay for what they use.

Cloud computing is a broad concept, not a specific technology or standard, different people from different perspectives will have different understanding, there is no authoritative definition.

1. The definition of cloud computing by analysts

   Early Merrill Lynch argued that cloud computing was the use of the Internet to run personal applications (E-mail, document processing, and presentations) and commercial applications (sales management, customer service, and financial management) on centrally managed servers. By sharing resources from these servers, such as storage and processing power, resources can be used more efficiently and costs can be reduced by 80% to 90%. Information Week, on the other hand, defines cloud computing more broadly: cloud computing is an environment in which any IT resource can be delivered as a service. The media is also interested in cloud computing. The Wall Street Journal, America’s best-selling magazine, is also keeping a close eye on the evolution of cloud computing. It argues that cloud computing enables enterprises to gain computing power, storage space, software applications, and data from very large data centers over the Internet. Customers pay only for the resources they use when necessary, avoiding the huge costs of building their own data centers and purchasing servers and storage devices.
2. The definition of cloud computing by enterprises

   IBM believes that cloud computing is a computing style based on the delivery of services, software, and processing power over public or private networks. Cloud computing focuses on the user experience, with the core separating the delivery of computing services from the underlying

technology. Cloud computing is also a way to share infrastructure, using pools of resources to connect public or private networks together to provide IT services to users. Eric Schmidt, Google's former CEO, argues that cloud computing distributes computing and data across a large number of distributed computers, making computing and storage capabilities highly scalable and allowing users to easily access applications and services over the network through a variety of access methods, such as computers and mobile phones. Its important feature is open, there will not be an enterprise can control and monopolize it. According to Kaifu Li, a former global vice president at Google, the entire Internet is a beautiful cloud where Internet users need to easily connect to any device, access any information, create content freely, and share it with friends. Cloud computing is based on open standards and services, the Internet as the center, to provide secure, fast, and convenient data storage and network computing services, so that the Internet "cloud" is to become every Netizen's data center and computing center. Cloud computing is actually Google's business model, and Google has been working hard to promote the concept.

Microsoft's approach to cloud computing is much more contradictory than Google's. If future computing power and software are all concentrated in the cloud, then clients don't need a lot of processing power, and Windows loses most of its power. As a result, Microsoft's approach has always been "cloud+end." Microsoft believes that the future of computing model is not just cloud computing. The "end" here refers to the client, which means that cloud computing must have a client to work with. "From an economic point of view, bandwidth, storage, and computing are not going to be free, and consumers need to find a model that fits what they need, so there must be end-of-the-line computing. In terms of communication supply and demand, although bandwidth has increased, content is also growing simultaneously, such as video and images. Bandwidth limitations are always there. From a technical point of view, the end of the computing power is strong, in order to bring users more exciting applications" said Dr. Yaqin Zhang, a former senior global vice president at Microsoft. Microsoft's definition of cloud computing is no different, it just underlines the importance of the "end" in cloud computing. Today, with the rise of Azure Cloud, Microsoft has embraced cloud computing across the scale.

The overview of cloud computing across the business market is shown in Fig. 1.5.

3. The definition of cloud computing by academia

In academia, Ian Foster, the father of grid computing, argues that cloud computing is a model of large-scale distributed computing driven by the economics of scale. In this model, abstract, virtualized, dynamically scalable, and managed computing power, storage, platforms, and services converge into a pool of resources that are delivered to external users on demand over the Internet. He

believes that several key points of cloud computing are: high scalability; can be encapsulated as an abstract entity and provide different levels of service for external users; economics resulting from scale; and services can be dynamically configured (via virtualization or other means) to deliver on demand.

Based on these different definitions, it's not hard to find out that the basic view of cloud computing is the same, but there are differences in the delimitation of certain areas. A more complete definition of cloud computing can be given from a comprehensive perspective: "Cloud computing is a computing model in which dynamically scalable and virtualized resources are delivered as services over the Internet." End-users don't need to know the details of the infrastructure in the cloud, do not need to have the appropriate expertise knowledge, do not need direct control, just pay attention to what resources they really need, and how to get the appropriate services over the network."

Zhu Jinzhi, who once worked at IBM, gave a relatively broad definition in his book *Smart Cloud Computing : The Platform of the Internet of Things* in order to cover cloud computing more comprehensively. The definition is as follows: "Cloud computing is a computing model: IT resources, data and applications are provided as services to users through the network." Its practical definition of "cloud" is a metaphorical method used to express the abstraction of complex infrastructure. Cloud computing is an abstraction of traditional computing infrastructure, so we choose to use "cloud" as a metaphor, as shown in Fig. 1.6.

Cloud computing starts with "software as a service," and then transforms all IT resources into services and provide to users. Think of cloud computing as a model that can easily access a common set of configurable computing resources (such as servers, storage devices, software, and services) through the network. These resources can be quickly provided and released, while minimizing management costs and the intervention of service providers.

We can look at cloud computing from two perspectives, the place where computing occurs and the form of resource supply. From the perspective of where computing occurs, cloud computing moves the operation of software from a personal computer (or desktop computer) to the cloud, that is, on a server or server cluster located in a "mysterious" geographic location. These servers or server clusters can be local, remote, or even far away. This seems to be a Client/Server (C/S) model, but cloud computing is not a traditional client/server model, but a huge improvement on this model. From the perspective of resource supply, cloud computing is a computing service, that is, all IT resources, including hardware, software, and architecture, are sold and charged as a service. For cloud computing, there are three main types of services: infrastructure as a service, providing hardware resources, similar to the traditional CPU, memory and I/O; platform as a service, providing an

environment for software operation, similar to traditional operating system and programming framework in programming mode; software as a service, providing application software functions, similar to application software in traditional mode. In the cloud computing model, users no longer purchase or buy out certain hardware, system software, or application software to become the owner of these resources, but purchase the usage time of the resource, and consume according to the billing model such as paying for the length of use.

It can be seen that cloud computing treats all resources as services and consumes them in a pay-as-you-go manner, which is the characteristic of the host era. In the host era, all users are connected to the host through a display terminal and a network cable, and billing is based on the consumed CPU time and storage capacity. The difference is that in the host mode, the calculation occurs on one host; in the cloud computing mode, the calculation occurs in a server cluster or data center.

Therefore, cloud computing is both a new computing model and a new business model. It is a new computing model because all computing is organized as a service; it is a new business model because the way users pay is different from the past, and pay according to what you use, which greatly reduces resource users' operating costs. It is not difficult to see that these two aspects of cloud computing rely on each other and are indispensable. Because only using resources as services can support the pay-as-you-go payment model; because the billing is based on what you use as you pay, resources can only be provided as services (not as packaged software or hardware). In fact, it can be said that cloud computing is a computing model, where computing boundary here is not determined by technical limitations, but by economic factors.

In a nutshell, cloud computing is the result of the hybrid evolution and integration of various concepts such as virtualization, utility computing, service computing, grid computing, and automatic computing. It started from mainframe computing and went through minicomputer computing, client /server computing, distributed computing, grid computing, and utility computing. It is not only a technological breakthrough (technical integration), but also a leap in business model (pay as much as you use, no waste). For users, cloud computing shields all the details of IT. Users do not need to have any knowledge or any control over the technical infrastructure of the services provided by the cloud, or even the system configuration and geographic location of the services provided. They only need to "turn on the switch"(Connect to the Internet) to enjoy the service.

It can be seen that cloud computing describes a new mode of supplying, consuming, and delivering IT services. This model is based on the Internet protocol and will inevitably involve the configuration of dynamically scalable and often virtualized resources. To some extent, cloud computing is a by-product of people's pursuit of easy access to remote computing resources.

The huge advantages of cloud computing in both technology and business model determine that it will become the leading technology and operating model of the IT industry in the future.

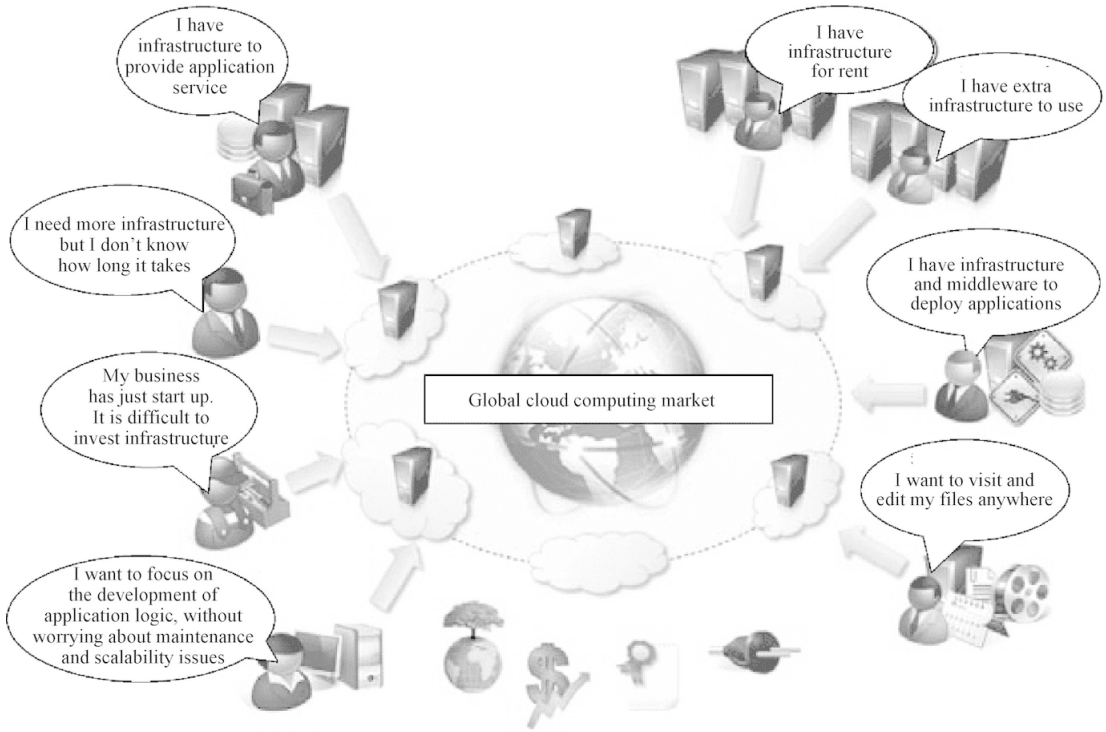

Fig. 1.5 Overview of cloud computing

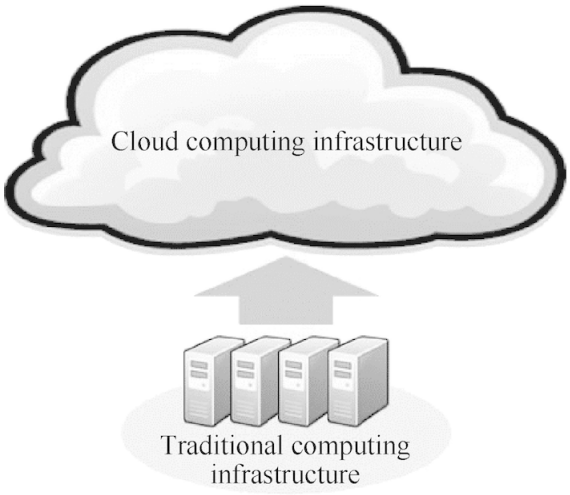

Fig. 1.6 Cloud computing is an abstraction of traditional computing infrastructure

1.4 The Emergence and Development of Cloud Computing

The origin of cloud computing can be traced back to the concept of utility computing proposed by John McCarthy in 1961. He mentioned: "If the kind of computer I am advocating becomes the computer of the future, then the computer may one day be organized into a public facility like a telephone system... Computer facilities may become the basis of a new and important industry." Contains the initial thoughts

of cloud computing. In the late 1990s, Salesforce took the lead in providing remote customized services to enterprises. In 2002, Amazon started operating the Amazon Web Services (AWS) platform, providing enterprises with services such as remotely customized storage, computing resources, and business functions. In 2006, Eric Schmidt, the CEO of Google, proposed the concept of “cloud computing.”

In fact, the emergence of cloud computing is not isolated, but the product of the development of computer technology and communication technology to a certain stage. Cloud computing technology is the product of a collection of various technologies.

There is a view that cloud computing is equivalent to the “Internet + computing” model, and the history of cloud computing is the history of the development of the Internet and computing models. So let’s briefly review the development history of the two.

1.4.1 The History of the Network and the Internet

In the early days, computers all operated on a stand-alone computer, and the calculation and transmission of data were all done on the local computer. The birth of the Internet connected these computers and ultimately connected the entire world. The following are some very representative milestones in the history of Internet development.

In 1969, ARPANET was born, and it is considered the predecessor of the Internet. There were only four nodes that first joined ARPANET, namely University of California, Los Angeles (UCLA); Stanford Research Institute (SRI); University of California, Santa Barbara (UC Santa Barbara); and University of Utah. The birth of ARPANET marked the beginning of the Internet era. In the following years, more and more nodes joined ARPANET, and more and more users in the non-military field. In 1983, out of security considerations, ARPANET separated 45 of these nodes to form a special military network called MILNET. The remaining nodes were used for civilian purposes.

In 1981, the first complete specification of TCP/IP was established, and the Internet has since had unified communication rules. TCP/IP is actually a collection of protocols, which includes Transmission Control Protocol (TCP), Internet Protocol (IP), and some other protocols. The earliest protocol used on ARPANET is called Network Control Protocol (NCP), but with the growth of ARPANET, NCP cannot meet the needs of large networks, while TCP/IP seems to be tailored for large or even giant network services. Therefore, in 1983, ARPANET replaced NCP with TCP/IP.

In 1983, ARPANET, PRNET, and SATNET were the three original networks that used TCP/IP communication. The first three networks switched to TCP/IP at the same time, marking the beginning of rapid development of the Internet.

In 1984, the Domain Name System (DNS) was born. After TCP/IP was adopted, the development of the Internet became more rapid, and more and more computers joined the network, and each computer used the digital IP address of the TCP/IP standard to identify each other. The common version 4 IP address (IPv4), an address corresponds to four bytes, and each byte is represented by a decimal number from 0 to 255. A typical IP address is in the form of 202.120.80.1. This type of digital IP address is not suitable for people to remember. It is like using an ID number to call the person you meet. I believe that few people can remember the ID number of everyone around them. Therefore, a new mechanism that is easy for people to remember is needed to replace the IP address to identify computers on the Internet. Thus, DNS came into being. DNS can realize the mutual conversion between digital IP addresses and domain names that are easier for people to remember, which is equivalent to using short, easy-to-remember names instead of ID numbers, which greatly reduces the difficulty of remembering. For example, the previous IP address is 202.120.80.1, and the associated domain name is www.ecnu.edu.cn, which represents the official website of East China Normal University. The domain name is hierarchical. For example, the official website of Huawei www.huawei.com, the com on the far right represents the enterprise or company, the Huawei in the middle represents the enterprise name, and the www on the far left represents the default website server name. For general corporate websites, there will also be a top-level domain name (cn is China) that represents the country (or region) on the far right. Through DNS, we can use the domain name to visit the corresponding homepage worldwide.

In 1986, the modern mail routing system MERS was developed.

In 1989, the first commercial network operator PSINet was established. Prior to the establishment of PSINet, most networks were funded by the government or the military for military, industrial, or scientific research. The establishment of PSINet represents that the Internet has entered the era of commercial operation.

In 1990, the first Internet search engine Archie appeared. In the early days of the Internet, although there was relatively little information on the Internet, there were already many valuable files (data), but these files were scattered on various File Transfer Protocol (FTP) servers, which made it difficult for users turn up. Therefore, a search engine or search website is needed for indexing and searching. So Archie was developed. Using Archie can easily find the location of the FTP server where the file is located by the file name, and then download it to the local with tools such as FTP.

In 1991, the World Wide Web (WWW) was invented by Tim Berners-Lee, a scientist at the European Center for Particle Research. It was a landmark technology in the history of the Internet. Spread and interconnected on the Internet. Hypermedia can be documents, voice or video, and it is a new way of expressing information. After the birth of the World Wide Web, some great and landmark

Internet companies were born, and various network applications that really changed people's lives began to emerge.

In 1995, e-commerce companies such as Amazon and eBay were established. In the history of Internet development, many companies have appeared, such as Yahoo! and Google. Amazon is the first Internet company to truly implement cloud computing. Amazon's early products sold were books. In order to process product information and user data, Amazon established a huge data center. In the United States, there is a "Black Friday" similar to "Double Eleven." On this day, Amazon needs to process a large amount of information, and all equipment in the data center will be turned on. But after this day, a lot of equipment will be idle. In order not to cause waste, Amazon will rent out the excess equipment. So in 2006, Amazon launched its first cloud computing product-EC2 (Elastic Compute Cloud).

In the late 1990s, the Internet was surging and experienced an "explosive" development. Domestic Internet companies "BAT" (Alibaba, Tencent, Baidu) were established during this period. The Internet has allowed people to see its magic, leading to rapid development, spawning a large number of "bubbles," and finally around 2000, the Internet bubble burst. However, after experiencing the bubble burst, the Internet has rapidly developed, and 2004 is also known as the "first year of social networking." With the rapid development of domestic Internet applications, from 2000 to 2020, a large number of domestic Internet companies have rapidly developed and become industry "giants." In addition to the BAT mentioned above, there are also JD.com, Ant Financial, ByteDance, Pinduoduo, Meituan, NetEase, Sina, Sohu, Didi, Suning, Xiaomi, etc. These Internet companies have penetrated deeply into people's daily lives. And even changed people's lifestyle to a certain extent.

1.4.2 The History of Computing Models

Cloud computing does not appear suddenly, but a result of the development and evolution of past technologies and computing models. It may not be the ultimate result of computing models, but a model suitable for current business needs and technical feasibility. The following describes the emergence of cloud computing by analyzing the development history of computing models. Figure 1.7 shows the development history of cloud computing from the perspective of computing models.

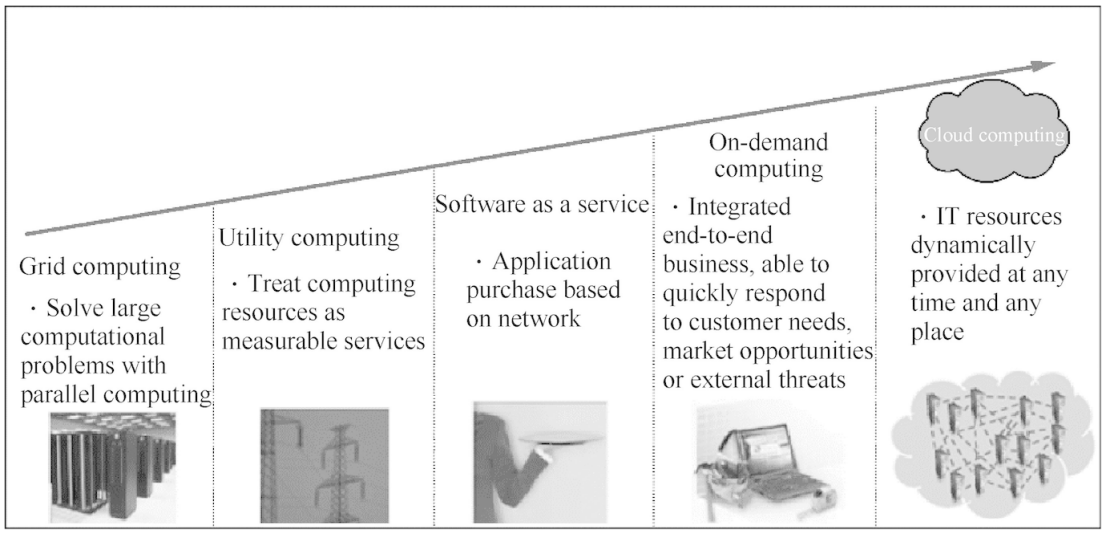

Fig. 1.7 Show the development history of cloud computing from the perspective of computing mode

Cloud computing is a master of multiple computing technologies and computing models. Therefore, it has similarities with many existing computing models. For example, the software runs in the cloud, and the way customers access cloud services through mobile terminals or clients is somewhat similar to the client/server model; the automatic scaling of cloud resources is somewhat similar to automatic computing; cloud computing gathers resources for customers to use, and grid computing, which was once a smash hit, is somewhat similar; a large number of computing nodes in cloud computing work at the same time, which seems to be somewhat similar to parallel computing; nodes that make up the cloud are distributed in multiple locations and are somewhat similar to distributed computing. Billing based on usage is somewhat similar to utility computing.

Although cloud computing does have similarities with various computing models that we are familiar with, they are not exactly the same. In fact, there are huge differences between cloud computing and certain computing models.

1. Cloud computing and mainframe computing

   As early as a few decades ago, shortly after the computer was invented, the computing model at that time had a “shadow” of cloud computing. In 1964, the world’s first mainframe System/360 was born, triggering a “revolution” in the computer and business fields.

   The market for the mainframe system is mainly enterprise users, and these users generally have a variety of business systems that need to use the mainframe system. So IBM invented the virtualization technology, which divides a physical server into many different partitions, and each partition runs an operating system or a set of business systems. In this way, each enterprise only needs to deploy a mainframe system to meet the needs of all business systems. Since the system has experienced decades of development, its stability is high and it has the reputation of “never shut

down.” The IBM mainframe is responsible for the most extensive and most important information and data processing tasks in the finance, communications, energy, and transportation industries. Before the emergence of cloud computing, more than 70% of the world’s corporate data was running on mainframe platforms, and most of the top companies in the world’s wealth rankings were using mainframes.

One characteristic of the mainframe is the concentration of resources, the concentration of computing and storage, which is a typical representative of the centralized computing model. Enterprises using mainframes will concentrate the business to be processed, usually in the form of batch processing and send them to the mainframe for processing. Most of the users of the mainframe use the terminal mode to connect to the mainframe, no data processing and storage are performed locally, and no measures such as patch management, firewall protection, and virus prevention are required. In fact, the mainframe system is the earliest “cloud, but these clouds are for specialized services, private networks, and specific fields.

Cloud computing and mainframe computing actually have many things in common, such as centralized management and billing based on usage. However, there is a big difference between cloud computing and mainframe computing. One of the important differences is the different user groups. The users of mainframe computing are usually large organizations and are prepared for key software, such as census, consumption statistics, enterprise resource planning (ERP), and financial transactions; while cloud computing is for the general public and can run a variety of large, medium, and small software. In addition, the processing unit of mainframe computing is usually a single mainframe, while the processing unit of cloud computing is generally composed of a large number of IT resources in a cluster manner, and the processing capacity is much greater than that of the early single mainframe.

2. Utility calculation

Utility computing emerged with the development of the mainframe. Considering the high purchase cost of the host, some users can only rent it instead of buying it. So some people put forward the concept of utility computing, whose goal is to package the server and storage system for users to use, and to charge users according to the amount of resources actually used by the users. This model is similar to the provision of water, electricity, gas, and telephone services, enabling users to use computer resources as if plugging a light bulb into a lamp holder. This model eliminates the need for users to own resources in order to use services and can also achieve the goal by leasing resources. Utility computing can be regarded as the predecessor of cloud computing.

The actual application of utility computing is mainly represented by IBM. IBM leases its own host resources to different users according to time. The host is still stored in IBM's data center, and users use IBM's resources remotely or on site in IBM's data center. The key technology in utility calculation is resource usage measurement, which guarantees the accuracy of pay-per-use.

From the perspective of the billing model, cloud computing is exactly the same as utility computing. Utility computing packages IT resources into measurable services for users to use, that is, CPU, memory, network bandwidth, and storage capacity are all treated as traditional utility usage (such as telephone networks) for packaging. The biggest advantage of this computing model is that users do not need to pay in advance, nor do they need to buy out IT resources. For most of the small- and medium-sized enterprises, there is not enough capital and technology to construct an IT infrastructure comparable to the Fortune 500 companies. They welcome the concept of utility computing because utility computing enables them to access and use advanced information technology and resources like the Fortune 500 like companies.

Compared with cloud computing, utility computing only specifies the billing model of IT assets and does not limit other aspects of IT assets, such as technology, management, configuration, and security. There are many more factors to consider in cloud computing, and the billing model is only one of the factors.

3. Client/server model

From the perspective of service access mode, cloud computing does have the shadow of a client /server model: customers connect with the remote cloud through a certain device and use the services provided by the application software running in the cloud. However, behind this similarity, the "remote server" provided by cloud computing has unlimited computing power, unlimited storage capacity, and never crashes, and all software can run on it. Users can also publish their own software to the "remote server," and the "remote server" can automatically configure the required resources for the software and change it as needed. In addition, cloud computing has its own set of models and rules (explained later), while the client/server model, on the other hand, refers to all Distributed Systems that can distinguish between a service provider (server) and a service requester (client).

4. Cluster computing

Since the cloud of cloud computing contains a large number of server clusters, it is very similar to cluster computing. However, cluster computing based on server clusters uses a tightly coupled group of computers to achieve a single purpose, while cloud computing can provide different supports according to the needs of users to achieve different purposes. In addition, cluster computing is limited distributed computing, which is not as complicated as the distributed computing faced by

cloud computing. In addition, cluster computing does not consider interactive end-users, while cloud computing does. Obviously, cloud computing includes elements of server cluster computing.

5. Service computing

The service provided by cloud computing is called cloud service, which is naturally reminiscent of service computing. Service computing is also called service-oriented computing and has the same concept as SaaS described later. This computing model provides all applications as services, and users or other applications use these services instead of buying out or owning software. In the service computing mode, different services are relatively independent, loosely coupled, and freely combined. For service computing, discovery of service is the key point.

Cloud computing has largely adopted the technology and way of thinking of service computing, but there are still important differences between service computing and cloud computing. First, although service computing is generally implemented on the Internet, service computing does not necessarily have to be provided in the cloud. A single server, small-scale server clusters, and a limited range of network platforms can provide service computing. Secondly, service computing is generally limited to providing services at the software level, while cloud computing extends the concept of services to hardware and operating environments, including the concepts of IaaS and PaaS. In other words, the concept of cloud computing is more extensive than the concept of traditional service computing.

6. Personal computer and desktop computing

In the 1980s, with the development of computer technology, the volume and cost of computer hardware were greatly reduced, making it possible for individuals to own their personal computers. The appearance of personal computers has greatly promoted the development of the software industry, and various end-consumer-oriented software has emerged. Software running on personal computers requires a simple and easy-to-use operating system. The Windows operating system just meets the needs of the public, and it has occupied the market with the popularity of personal computers. Personal computers have their own independent storage space and processing capabilities. Although their performance is limited, they are sufficient for individual users within a period of time. Personal computers can complete most of the personal computing needs; this model is also called desktop computing.

Before the advent of the Internet, the sales model of software and operating systems was an authorization model, that is, the software code was copied to a computer through a floppy disk or CD-ROM, and each copy required a payment to the software developer. After several years of development of this model, some problems appeared, such as high cost and cumbersome software

upgrades. The purpose of the upgrade is to solve some of the previous problems or to use new features, but the upgrade process can sometimes be cumbersome. For a large enterprise, its IT department may need to manage hundreds of software, thousands of versions, and tens of thousands of computers. Each version of the software needs to be maintained, including problem tracking, patch management, version upgrades, and data backup. This is by no means a simple job.

7. Distributed computing

The personal computer did not solve the problem of data sharing and information exchange, so the network-Local Area Network (LAN) and later the Internet appeared. The network connects a large number of computers distributed in different geographical locations, including personal computers and servers (large mainframes and later medium and small mainframes). With so much computing power, can an application run on multiple computers to complete a computing task together? The answer is of course yes, this is distributed computing.

Distributed computing relies on distributed systems. A distributed system consists of multiple computers connected through a network. Each computer has its own processor and memory. These computers cooperate with each other to complete a goal or computing task together. Distributed computing is a large category, which includes many familiar computing modes and technologies, such as grid computing, P2P computing, client/server computing, and Browser/Server (B/S) computing. Of course, cloud computing is also included. In today's network age, there are very few non-distributed computing applications, and only some stand-alone applications fall into this category, such as word processing and stand-alone games.

8. Grid computing

One of the main functions of computers is to perform complex scientific calculations, and a "master" in this field is supercomputers, such as China's "Galaxy" series, "Dawn" series, "Tianhe" series, and "Shenwei·Light of Taihu Lake." In foreign countries, there are Japan's Fugaku, which ranks first in the global supercomputing rankings in 2020, and the US's Summit, which is the second. The computing model centered on supercomputers has obvious shortcomings: Although it is a "big Mac" with powerful processing capabilities, it is extremely expensive, and is usually only used by some state-level departments (such as aerospace, meteorology, and military industries) who have the ability to configure such equipment. As people increasingly need computers with more powerful data processing capabilities, people began to look for a low-cost computing model with superior data processing capabilities. Finally, scientists found the answer, that is, grid computing.

Grid computing appeared in the 1990s. It is a new type of computing model that has been developed rapidly with the development of the Internet, specifically for complex scientific

computing. This computing model uses the Internet to organize computers distributed in different geographical locations into a “virtual supercomputer.” Each computer participating in the calculation is a “node,” and the entire calculation is composed of thousands of “nodes,” so this calculation mode is called grid computing. In order to perform a calculation, grid computing first divides the data to be calculated into a number of “small pieces,” and then distributes these small pieces to each computer. Each computer executes its assigned task segment and returns the calculation result to the master control node of the calculation task after the task calculation is completed.

It can be said that grid computing is an extension of supercomputers and cluster computers. Its core is still trying to solve a huge single computing problem, which limits its application scenarios. In fact, in non-scientific fields, only a limited number of users need to use huge computing resources. Grid computing once became “hot” after entering the twenty-first century. Major IT companies have made many investments and attempts, but they have not found a suitable use scenario. In the end, grid computing has made a lot of progress in the academic field, including some standards and software platforms have been developed, but it has not been popularized in the commercial field.

To some extent, many things that grid computing must do are also things that cloud computing must do, but grid computing cannot be regarded as cloud computing. First of all, grid computing is mainly for scientific computing and simulation, while cloud computing is made for general public. Secondly, grid computing does not consider interactive end-users, while cloud computing must consider.

9. SaaS

SaaS is a computing model that uses software as a service. It is a mode of providing software through the Internet. Service providers uniformly deploy the software on their own servers. Users can order the required software services from the service provider through the Internet according to their actual needs and pay the service provider according to the number of services ordered and the length of time and obtain the services provided by the service provider through the Internet. Users no longer need to purchase software, but instead rent Web-based software from service providers to manage business activities, and there is no need to maintain the software. The service provider will have full authority to manage and maintain the software. While providing Internet software to users, service providers also provide offline operation of the software and local data storage, so that users can use the software and services they have ordered anytime, anywhere.

SaaS first appeared in 2000. At that time, with the vigorous development of the Internet, various new business models based on the Internet continued to emerge. For traditional software companies, SaaS is the most significant change. This model turns the one-time software purchase income into continuous service income. Software providers no longer calculate how much software they sell but need to always pay attention to how many paying users there are. Therefore, software providers will pay close attention to their own service quality, and continuously improve their own service functions to enhance their own competitiveness. This model can reduce piracy and protect intellectual property rights, because all the code is at the service provider, users cannot obtain it, nor can it be cracked or decompiled.

In summary, in addition to utility computing, the computing models discussed above are all technical aspects, and the cloud computing model covers both technical and commercial aspects. This may be the biggest difference between cloud computing and the above-mentioned computing models.

10. The emergence of cloud computing

Looking back at the history of the development of computing models, it can be summarized as: Concentration-Decentralization-Concentration. In the early days, limited by technical conditions and cost factors, only a few companies could have computing power. Obviously, the computing model at that time could only be centralized. Later, with the miniaturization and low cost of computers, computing also became decentralized. Up to now, there is a trend toward centralization in computing, and this is cloud computing.

Users can use cloud computing to do many things. As we can see from the previous article, there are three basic services provided by cloud computing: the first one is hardware resource services; the second is operating environment services; and the third is software services. Then users can also use the cloud computing platform (also called cloud platform) in at least three ways: the first one is to use the cloud platform to save data (using the hardware resources provided by the cloud environment); the second is to run software on the cloud platform (using the cloud the operating environment of the environment); the third is to use software services on the cloud platform (using software services arranged on the cloud, such as maps, search, and mail).

The service provided by cloud computing is not only the IT resource itself, if it is nothing more than that, there is no need to develop cloud computing. Storage of data, running programs, and using software can be implemented on many platforms, and cloud computing is not required. The reason for using cloud computing is a result of the way and the ability to provide resources. Cloud computing has great advantages in resource provision methods and capabilities.

In addition to the scale advantages of the cloud platform mentioned above, another important advantage of cloud computing is elastic resource allocation. The resources provided by cloud services are more when need is high, and less when need is low. If we deploy an application software on the cloud, the cloud controller will dynamically adjust the resources allocated to the application according to the changes in customer needs of the software, so as to ensure that it can meet any sudden increase in customer demand at any time and at the same time avoid waste of resources when customer demand is low.

In addition, cloud platform also provides another advantage that most people may not know. Some operations can only have a powerful effect if they are placed in the cloud, while direct deployment on the business location or the user's client has limited or no effect. This is because technically, the desktop or server operation mode is no longer adequate for many challenges faced by IT systems, and these challenges can be solved in the cloud. For example, in terms of virus checking and killing, the antivirus software on the desktop has a lackluster checking and killing effect, and the highest antivirus efficiency can only reach 49% to 88% (according to the data of Arbor Networks in the United States). In addition, scanning and killing also occupies a large amount of computing resources of the personal computer, resulting in extremely low efficiency of the entire system. But moving the antivirus operation to the cloud can solve this problem. Note that moving antivirus operations to the cloud mentioned here is very different from the "cloud antivirus" promoted by some antivirus software companies. The cloud antivirus technology on the market refers to deploying antivirus software in the cloud and checking and killing remote clients through the network. The only advantage of this kind of cloud antivirus is that it is easier to update and maintain antivirus software, but the antivirus capability has not improved. However, if a variety of different antivirus software is deployed in the cloud and the network data is cross-checked and killed in the cloud, the antivirus efficiency can be increased to more than 96%, and it will not occupy the computing resources of the client.

Another example is, in the market, the operating modes of personal computers, servers, or clusters face difficulties in updating and maintaining. Only changing the functional role of a server (replacement of the system and software) requires a lot of effort and is prone to errors. The installation, configuration, upgrade, and other management operations of software distributed on various computers in the organization are a headache for many companies. Moving these services to the cloud can solve these problems.

In general, cloud computing mainly has the following four advantages:

- On demand, unlimited computing resources.

- Instantly available software/hardware resources and IT architecture.
- Charge by usage model.
- Processing environment that is difficult to provide by a single machine.

Although various service concepts with the title “cloud computing“emerge in an endless stream, not every service can be classified as a cloud computing service. How to judge whether a service is a true cloud computing service? Generally speaking, you should see whether the following three conditions are satisfied at the same time.

1. The service should be accessible anytime, anywhere. Users can use the services at anytime, anywhere, and through any device that can connect to the Internet, without having to consider the installation of applications or the implementation details of these services.
2. The service should always be online. Occasional problems may occur, but a true cloud computing service should always ensure its availability and reliability. That is, ensure that it can be accessed through the network at any time and provide services normally.
3. The service has a large enough user base. This is the so-called “multi-leasing,” a “leasing” in which a basic platform provides services to multiple users. Although there is no clear number to divide, it is only for a small number of users, that is, using cloud computing-related technologies to support its basic system architecture, it should not be classified as a cloud computing service. Because only a large user base will generate pressure to access massive amounts of data. This is the most fundamental reason for the emergence of cloud computing, and it is also one of the signs that cloud computing services are different from other Internet services.

1.4.3 The Driving Force of Cloud Computing

Cloud computing does not appear from nowhere. Its emergence is promoted by a variety of factors, which has a certain inevitability.

1. The improvement of internet broadband

   Bandwidth is a necessary condition for the popularization of cloud computing. Since computing and storage are placed on the other side of the network, it is necessary to allow users to easily access these data. In recent years, with the popularization of the Internet, major network operators have also continued to improve their Internet infrastructure. On the one hand, the bandwidth of the core network is rapidly expanding; on the other hand, the network access of home and business users has also undergone essential changes. Take home users as an example. From the very beginning dial-up Internet access (speeds in the range of tens to hundreds of kilobits per second), to the later

Asymmetric Digital Subscriber Line (ADSL) (speeds in the hundreds of thousands). Bits per second to several megabits per second), and then to the current fiber to the home (network speeds are tens to hundreds of megabits per second or even higher). The increase in bandwidth has changed the mode of use of the network and the types of network application. With the development of 4G/5G technology, network bandwidth will further increase until users do not perceive the limitation of bandwidth.

2. Technology maturity

   There are many similarities between cloud computing and utility computing, but utility computing is not really popular because of the lack of sufficient operability. Any idea, if there is no practical way to realize it, will become a fantasy. The recognition of cloud computing by the public is also closely related to its technological maturity. Cloud computing corresponds to not one technology, but a combination of multiple technologies, which turns the concept of IT as a service into reality. At different levels, different technologies may be used.

   These technologies are hidden in the background and are invisible to users. This is also the hidden part of the cloud. We can imagine a data center that provides cloud computing services as a huge factory filled with hundreds of servers and connected by intricate cables. Many intelligent applications are running on these servers. They can manage these servers efficiently, ensure that the system can automatically recover when server fails, and also ensure that the entire center is running at a very low cost.

3. The Development of Mobile Internet

   The rapid development of the mobile Internet has led to a rapid increase in the number of digital mobile terminal devices represented by mobile phones and tablet computers. On average, everyone in the country has multiple digital mobile terminal devices that can access the Internet. How to manage the data in these devices has become a big problem. One is that these devices cannot all have strong computing power; the other is that data is scattered on each device, with duplication and redundancy, and the same data may also exist in both new and old versions. So the cloud computing model has become an ideal solution to this problem. For example, users may need to uniformly manage photos on computers, mobile phones, and digital cameras. Although they can be copied or transferred between devices, it is very troublesome. If these devices are connected to the Internet, the photos are synchronized to the cloud via the Internet, and the cloud-based photo management is performed, and the classification, update, synchronization, and access of the photos become very convenient.

4. The evolution of data center

   For users, a data center is a “factory” that provides computing and storage capabilities on the other end of the Internet and is a “power plant” for the IT industry. Data centers are unfamiliar to

ordinary Internet users, just as everyone who uses electricity does not care about how power plants operate. In fact, data centers are constantly evolving. Data centers can be divided into two types: one is to provide services to the Internet; the other is private to the enterprise and only open to the inside. Regardless of the type, the data center needs someone to operate it to ensure that it can provide services uninterrupted. According to a survey, more than 90% of 1000 organizations worldwide believe that they need to make major changes to their data centers in recent years. For them, the current challenges include expensive management costs, rapidly increasing energy consumption, rapidly increasing user demand, and inefficient use of IT resources. In view of these problems, data centers urgently need a new architecture and management concept, and cloud computing is a solution from the perspective of service providers.

5. Economic factor

When a product is technically feasible and has a wide range of needs, the only factor that determines its success or failure is the price, or user cost. The most fundamental factor in changing the computing mode is cost, and technology is the triggering condition. In the era of mainframes, the main reason for the use of centralized computing is that the cost of the mainframe is too high, and the appearance of personal computers has greatly reduced the user's cost of use, so that each enterprise can afford its own data center at a price. Today, the emergence of the Internet and cloud computing has made it possible to further reduce costs. If costs can be reduced, companies will of course consider adopting new technologies.

What is the trick to saving costs in cloud computing? In fact, it is the scale effect. For example, for power generation, each household uses its own generator to generate electricity. Obviously, the total cost is higher than that of centralized power supply through power plants. Another example is transportation. It is obviously more economical to use a bus to transport the same number of people than a car. Through scale, cloud computing can not only reduce fixed asset investment, but also reduce operating costs. When resources are concentrated, time-sharing or partition sharing of resources can make the same resources play a greater role. Coupled with intelligent resource allocation, the maximum use of resources can be realized. As far as energy usage efficiency is concerned, the value of Power Usage Effectiveness (PUE) has become an internationally accepted measure of data center power usage efficiency. The PUE value refers to the ratio of all energy consumed by the data center to the energy consumed by the IT load. The benchmark is 2, and the closer to 1, the better the energy efficiency level or the higher the power usage efficiency. The average PUE value of the data center is 1.21, so the use of the data center can greatly save energy.

6. Big data

Big data is another major driving force for cloud computing. Because processing massive amounts of data requires massive storage capacity and massive computing power. The general IT architecture is already incompetent, so standard equipment clusters were born, which evolved into cloud computing platforms. In fact, the two well-known commercial cloud platforms-Amazon's AWS and Google's App Engine are both spawned by processing big data.

In addition, some other driving forces that promote the development of cloud computing include the following.

- Improve resource utilization, save energy, and reduce consumption: Cloud computing (strictly speaking, virtualization) can increase server utilization from 15% to 60% or even higher, thereby reducing the energy consumption of unit computing tasks.
- Reduce the maintenance cost of the information system: The maintenance is all in one place and completed by specialized personnel.
- Improve the security posture of IT assets: All security issues are solved in one place, which is much easier than scattered in the business location of many users.
- Improve the disaster recovery capability of the information system: Cloud computing providers can conduct centralized investment and management for disaster recovery.

All in all, the driving force that promotes the emergence and development of cloud computing is economic, flexibility, convenience, elasticity, unlimited, and charge by usage.

1.4.4 The Development of Cloud Computing

Since the initial appearance of the concept of cloud computing, enterprise IT architecture has evolved from a traditional non-cloud architecture to a target cloud-based architecture. In summary, it has experienced the following three major milestone development stages.

1. Could computing 1.0

   IT infrastructure resource virtualization stage for data center administrators. The key feature of this stage is that through the introduction of computing virtualization technology, enterprise IT applications are completely separated and decoupled from the underlying infrastructure, and multiple enterprise IT application instances and operating environments (guest operating systems) are reused in on the same physical server. And through virtualized cluster scheduling software, more IT applications are reused on fewer server nodes, thereby achieving an improvement in resource utilization efficiency.
2. Cloud computing 2.0

Resource servicing and management automation stage for infrastructure cloud tenants and cloud users. The key features of this stage are reflected in the introduction of standardized services and resource scheduling automation software on the management plane, as well as software-defined storage and software-defined network technologies on the data plane, for internal and external tenants, This would transform the complex and inefficient application, release, and configuration process of infrastructure resources that originally required manual intervention by data center administrators into one-click, fully automated resource distribution service process under necessary restricted conditions (such as resource quotas and permission approval). This change has greatly improved the rapid and agile distribution of infrastructure resources required for enterprise IT applications, shortened the preparation cycle of infrastructure resources required for enterprise IT applications to go online, and transformed the static rolling plan of enterprise infrastructure into the elastic on-demand supply of dynamic resources. This change also laid the foundation for enterprise IT to support its core business to move toward agility and better respond to the ever-changing business competition and development environment of the enterprise. In the cloud computing 2.0 stage, the provision of infrastructure resource services for cloud tenants can be in the form of a Virtual Machine (VN), a container (lightweight virtual machine), or a physical machine. The evolution of enterprise IT cloudification at this stage does not involve changes in enterprise IT applications, middleware, and database software architectures above the infrastructure layer.

3. Cloud computing 3.0

Distributed microservices of enterprise application architecture for enterprise IT application developers and management and maintainers, Internet reconstruction of enterprise data architecture, and big data intelligence stage. The key feature of this stage is reflected in the fact that the enterprise IT’s own application architecture has gradually shifted from (relying on traditional business databases and middleware business suites, specifically designed for each business application field, chimney-like, high-complexity, stateful, large-scale) vertical scale application layered architecture to (relying on open source enhanced, highly shared across different business application domains) database, middleware platform service layer and (more lightweight and decoupling functions, complete separation of data and application logic) distributed stateless architecture. This enables enterprise IT to reach a new level in supporting enterprise business agility, intelligence, and resource utilization efficiency and pave the way for the rapid iterative development of enterprise innovative business.

Regarding the above three development milestones, cloud computing 1.0 is already the past, and some industries and enterprise customers have completed the initial scale of cloud computing 2.0 construction and commercial use and are considering further expansion at this stage and the evolution

toward cloud computing 3.0. Another part of the customers is moving from cloud computing 1.0 to cloud computing 2.0, and even start the evaluation and implementation of the evolution of cloud computing 2.0 and cloud computing 3.0 simultaneously.

1.5 The Advantage of Cloud Computing

The implementation and innovation of any technology is to meet the application needs of a certain group of people. Cloud computing is not an exception. It gradually penetrates into all areas of people's life and production, bringing convenience and benefits to people. The advantages of cloud computing are as follows:

1. Cut costs

   Through cloud computing, companies can minimize or completely cut initial investment because they do not need to build data centers or build software/hardware platforms on their own, nor do they need to hire professionals for development, operation, and maintenance. It is usually much cheaper to use cloud computing services than to purchase software/hardware to build the required system.
2. Data can be accessed instantly anytime, anywhere

   "Cloud" brings greater flexibility and mobility. Using the cloud, companies can instantly access their accounts through any device anytime, anywhere; data can be stored, downloaded, restored, or processed easily, saving a lot of time and effort.
3. Improve adaptability and flexibly expand it needs

   In most cases, the capacity of the IT system does not match the needs of the enterprise. If an enterprise configures IT equipment according to the peak demand, it will be idle at ordinary times, resulting in a waste of investment. If an enterprise configures IT equipment according to average demand, it will not be enough during peak demand. However, with cloud services, companies can have more flexible choices and can increase, decrease, or release the resources they apply for at any time.
4. Unified platform

   Companies may be running different types of platforms and devices at the same time. In the cloud service platform, the application and the hardware platform are not directly related, thereby eliminating the need for multiple versions of the same application.

1.6 Classification of Cloud Computing

The layering of clouds focuses on the construction and structure of the cloud, but not all clouds of the same construction are used for the same purpose. Traditional operating systems can be divided into desktop operating systems, host operating systems, server operating systems, and mobile operating systems. Cloud platforms can also be divided into many different types. Cloud classification is mainly based on the cloud's operating mode and service mode. The former category is concerned with who owns the cloud platform, who is operating the cloud platform, and who can use the cloud platform. From this perspective, clouds can be divided into public clouds, private clouds (or dedicated clouds), community clouds, hybrid clouds, and industry clouds. The latter classification is based on the service model of cloud computing, and the cloud can be divided into three layers: IaaS, PaaS, and SaaS.

1.6.1 Classification by Operating Model

1. Public cloud

   Public cloud is a type of cloud environment that can be publicly accessed, usually owned by a third-party cloud service provider. It is called public cloud because it can be accessed by the unrestricted public. Public cloud service providers can provide the installation, management, deployment, and maintenance of IT resources in all aspects, from applications and software operating environments to physical infrastructure. End-users achieve their goals through shared IT resources, and only pay for the resources they use, and obtain the IT resource services they need in this relatively economical way.

   In the public cloud, users do not know with whom to share resources, and how the underlying resources are implemented, and they cannot control the physical infrastructure. Therefore, the cloud service provider must guarantee the security and reliability of the provided resources and other non-functional requirements. The level of these non-functional services also determines the service level of the cloud service provider. For those cloud services that need to strictly comply with security and regulatory compliance, higher level and more mature service levels are required. Examples of public clouds include foreign Google App Engine, Amazon EC2, IBM Developer, etc. domestic Tencent Cloud, Alibaba Cloud, Huawei Cloud, Ucloud, etc.

2. Private cloud

   Enterprises and other social organizations are not open to the public. Data centers that provide cloud services (IT resources) for the enterprises or organizations are called private clouds. Compared with public clouds, users of private clouds own the entire cloud center facility, can control where program run and can decide which users are allowed to use cloud services. Since private cloud services are provided for enterprises or organizations, private cloud services can be less subject to

many restrictions that must be considered in public clouds, such as bandwidth, security, and regulatory compliance. Moreover, private clouds can provide more guarantees of security and privacy through means such as user range control and network restrictions.

The types of services provided by private clouds can also be diversified. Private cloud can not only provide IT infrastructure services, but also support cloud services such as application and middleware operating environment, such as internal management information system (IMS) cloud services.

3. Community cloud

Both public and private clouds have disadvantages. A compromised cloud is the community cloud. As the name suggests, it is a cloud platform owned by a community, not an enterprise. Community cloud generally belongs to a certain enterprise group, institution alliance or industry association, and generally also serves the same group, alliance, or association. If some organizations are closely connected or have common (or similar) IT needs and trust each other, they can jointly construct and operate a community cloud in order to share infrastructure and enjoy the benefits of cloud computing. All members of the group can use the community cloud. In order to facilitate management, community cloud is generally operated and maintained by one organization, but it can also be managed by a cloud platform operation and maintenance team formed by multiple organizations.

Public cloud, private cloud, and community cloud are shown in Fig. 1.8.

4. Hybrid cloud

Hybrid cloud combines “public cloud” and “private cloud” together. Users can partly own and share partly with others in a controlled way. Enterprises can take advantage of the cost advantages of public clouds to run non-critical applications on the public cloud, and at the same time provide services through the internal private cloud for major applications with higher security requirements and more criticality.

There are many reasons for using hybrid cloud. There are two main reasons: the compromise of various considerations; the transition from private cloud to public cloud. For the first reason, although some organizations are eager to use the public cloud, because of various regulations, confidentiality requirements or security restrictions, they cannot put all their resources on the public cloud, so some IT resources will be deployed in the public cloud. In the above situation, part of the IT resources is deployed in the business location, which will form a hybrid cloud.

In the long run, public cloud is the mainstream of cloud computing development due to its higher resource utilization efficiency, but private cloud and public cloud will coexist for a long time in the

form of common development. Just like the emergence of banking services, the transfer of currency from individuals to bank custody is a safer and more convenient process, but some people may choose to keep them on their own.

5. Industry cloud

   The industry cloud is for the purpose of the cloud, not for the owner or user of the cloud. If the cloud platform is customized for a certain industry (e.g., for the automotive industry), it is called an industry cloud. The components used in the industry cloud ecological environment should be more suitable for related industries, and the software deployed on it is also industry software or its supporting software. For example, for the cloud platform established by the hospital, the data storage mechanism deployed above should be particularly suitable for the storage, indexing, and query of medical data.

   There is no doubt that the industry cloud is suitable for the specified industry, but it may be of little value to the average user. Generally speaking, the structure of the industry cloud will be simpler, and its management is usually taken care of by the industry's "leading" or a computing center (supercomputer center) designated by the government.

   The relationship between the industry cloud and the four types of clouds mentioned above is not exclusive, and there may be an overlapping or overlapping relationship between them. For example, industry clouds can be built on public clouds, private clouds, and more likely community clouds.

6. Other cloud types

   In addition to the cloud types above, there are other cloud types. For example, according to whether the cloud is aimed at individuals or enterprises, it can be divided into consumer cloud and enterprise cloud. The consumer cloud audience is the general public or individuals, so it is also called the public cloud. This kind of cloud promotes personal storage and document management needs; the enterprise cloud is for enterprises and promotes comprehensive IT services for enterprises. The classification of these clouds is still a certain segmentation or combination of the above cloud types in essence.

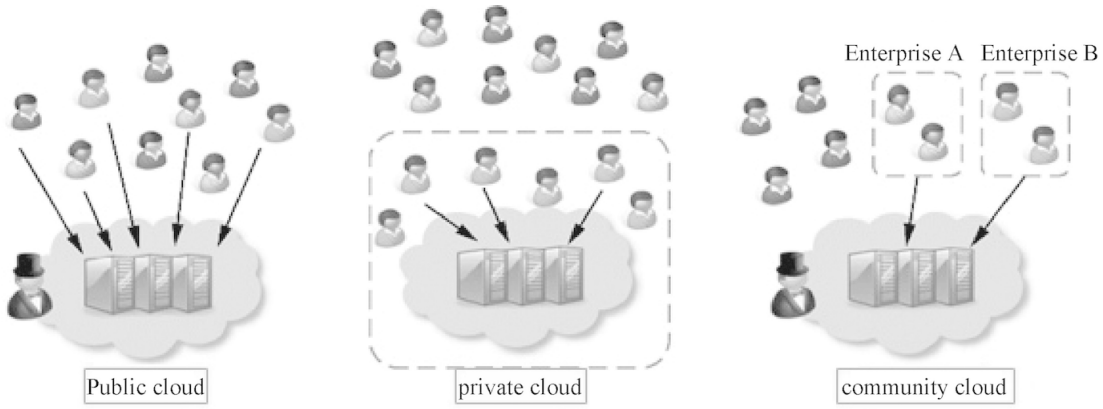

Fig. 1.8 Public cloud, private cloud, and community cloud

1.6.2 Classification by Service Model

According to the service model of cloud computing, the cloud can also be divided into three layers: IaaS, PaaS, and SaaS. Different cloud layers provide different cloud services. Figure 1.9 shows the composition of a typical cloud computing.

1. IaaS

   IaaS is at the bottom of the three-layer service of cloud computing, and it is also the scope covered by the narrow definition of cloud computing. IaaS provides IT infrastructure to users in the form of services like water and electricity and provides highly scalable and on-demand IT capabilities based on hardware resources such as servers and storage in the form of services. It is usually charged according to the cost of the resources consumed.

   This layer provides basic computing and storage capabilities. Taking the provision of computing capabilities as an example, the basic unit it provides is a virtual server, including CPU, memory, operating system, and some software, as shown in Fig. 1.10. Specific instance is Amazon EC2.

2. PaaS

   PaaS is located in the middle of the three-layer service of cloud computing and is often referred to as a “cloud operating system,” as shown in Fig. 1.11. It provides end-users with an Internet-based application development environment, including application programming interfaces and operating platforms and supports various software/hardware resources and tools required for the entire life cycle of applications from creation to operation. The billing is usually based on user or login status. At the PaaS layer, service providers provide encapsulated IT capabilities, or some logical resources, such as databases, file systems, and application operating environments. Examples of PaaS products include Huawei’s software development cloud DevCloud, Salesforce’s Force.com, and Google’s Google App Engine.

PaaS is mainly for software developers. It used to be a difficult problem for developers to write and run programs in a cloud computing environment through the network. Under the premise of gradual increase in network bandwidth, the emergence of two technologies has solved this problem. One is online development tools. Developers can use browsers, remote consoles (running development tools in the console), and other technologies to directly develop applications remotely, without the need to install development tools locally; the other is local development tools and cloud computing integrated technology, that is, deploying the developed application to the cloud computing environment through local development tools, while enabling remote debugging.

3. SaaS

SaaS is the most common cloud computing service, located at the top of the three-tier cloud computing service, as shown in Fig. 1.12. The user uses the software on the Internet through a standard Web browser. Cloud service providers are responsible for maintaining and managing software and hardware facilities and provide services to end-users for free or on-demand rental.

These services are both for general users, such as Google Calendar and Gmail, and for enterprise groups to help with payroll processes, human resource management, collaboration, customer relationship management and business partner relationship management, such as Salesforce.com and Sugar CRM. These SaaS-provided applications reduce the time for users to install and maintain software and their skills requirements and can reduce software license fees through pay-per-use.

Fig. 1.9 Components of cloud computing

Fig. 1.10 IaaS structure

Fig. 1.11 PaaS structure

Fig. 1.12 SaaS structure

The above three layers, each has corresponding technical support to provide the services of this layer, with the characteristics of cloud computing, such as elastic scaling and automatic deployment. Each layer of cloud services can be independent into a cloud or based on the services provided by the clouds below. Each kind of cloud can be directly provided to end-users for use, or it can only be used to support upper-layer services. The three types of service models usually have different user groups (see Fig. 1.13).

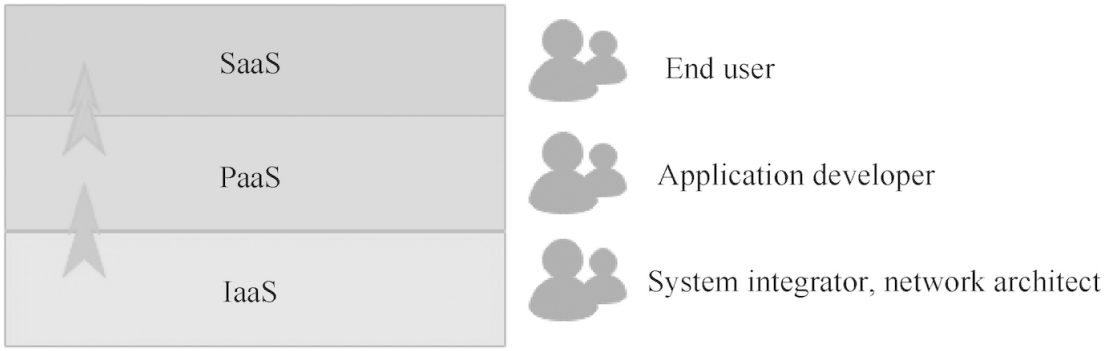

Fig. 1.13 Types of user groups of the three types of service models

1.7 Cloud Enabling Technology

This section introduces some basic key technologies included in the current cloud computing technology, which called cloud enabling technology. This includes broadband network and internet architecture, data center technology, virtualization technology, Web technology, multi-tenant technology, and service technology.

From a technical point of view, cloud computing is inextricably linked with various technologies such as distributed systems, virtualization technology, and load balancing. Like the “Je Kune Do” in information technology although it combines the essence of various types, it still forms its own types.

In terms of specific technical realization, the cloud platform innovatively integrates a variety of technical ideas, through different combinations, to solve different problems encountered in specific applications. Therefore, people will find a variety of technologies in the cloud platform, and some people will also judge that cloud computing is nothing but an old tune. However, if we only focus on the existence of a certain technology and ignore the integration and innovation of cloud computing itself in technical applications, there will be a situation of “seeing the trees, not the forest,” which is not only biased, but also leads to perception errors.

As far as technology is concerned, cloud computing is essentially derived from ultra-large-scale distributed computing and is an evolved distributed computing technology. Cloud computing also extends the Service-Oriented Architecture (SOA) concept and integrates virtualization, load balancing, and other technologies to form a new set of technical concepts and implementation mechanisms. Specifically, the core significance of cloud computing lies not only in the development of technology, but also in the organization of various technologies to change people’s thinking about building IT systems and at the same time make fundamental changes in the structure.

1.7.1 Broadband Network and Internet Architecture

All clouds must be connected to the network, and this inevitable requirement forms an inherent dependence on network interconnection. The Internet allows remote provision of IT resources and directly supports ubiquitous network access. Although most clouds rely on the Internet for access, cloud users can also choose to access the cloud only through private or proprietary network connections. The attractiveness of the cloud platform is closely related to the quality of the service provided by the access network.

The Internet's largest backbone network is established and deployed by ISPs, and they rely on core routers for strategic interconnection. These routers are in turn connected to transnational networks in the world. Figure 1.14 shows the interconnection between a backbone ISP network and other ISP networks and various organizations.

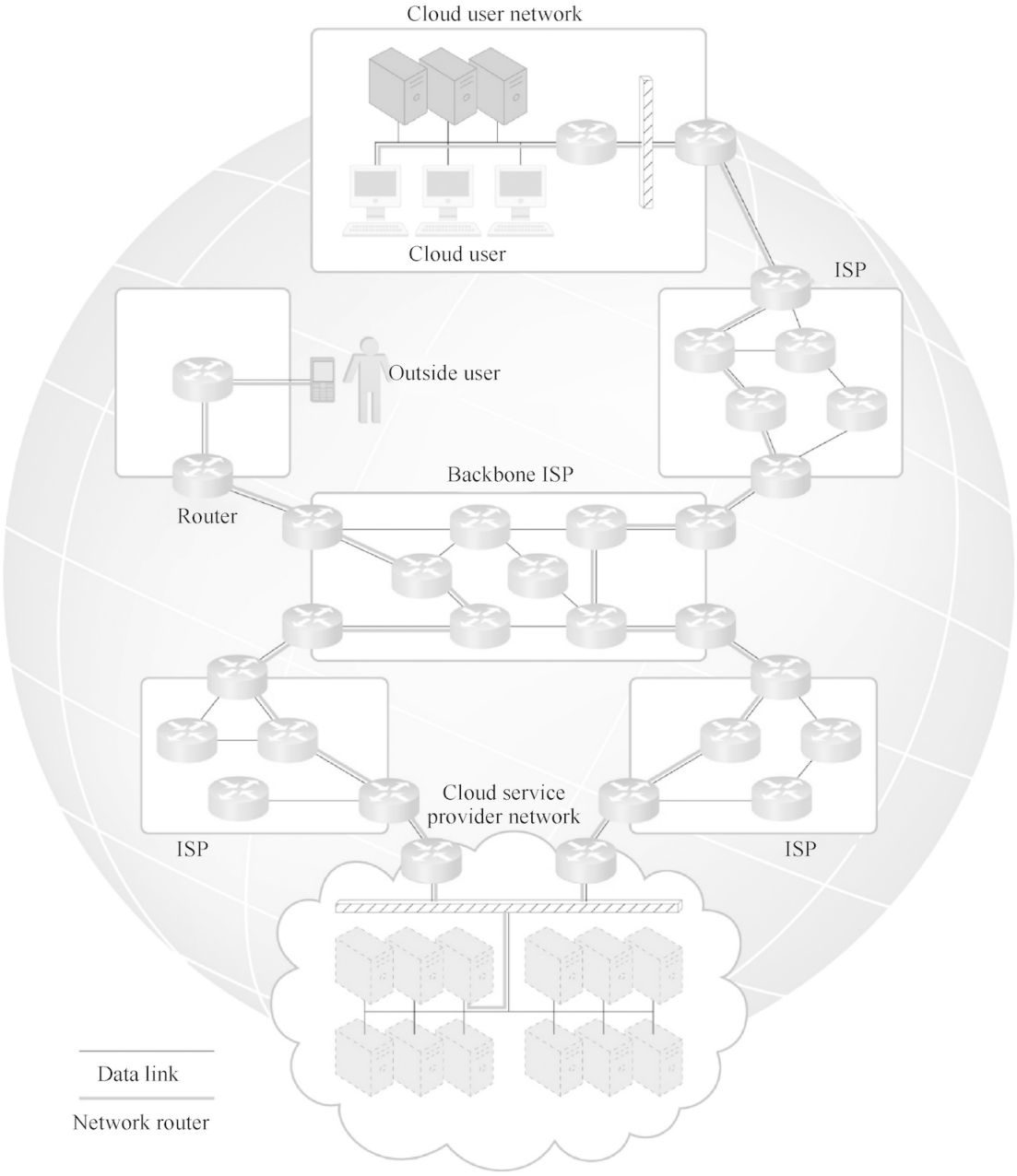

Fig. 1.14 The network of a backbone ISP is interconnected with the networks of other ISPs and various organizations

The concept of the Internet is based on a non-centralized supply and management model. ISPs can freely deploy, operate, and manage their networks, and they can also freely choose other ISPs that need to be interconnected. Although there are companies such as The Internet Corporation for Assigned Names and Numbers (ICANN) that supervise and coordinate important Internet affairs, there is actually no central entity to fully control the Internet. For specific countries, government regulations and regulatory laws supervise the services provided by domestic institutions and ISPs.

The topology of the Internet has become a dynamic and complex collection of ISPs. These ISPs are highly interconnected through their core protocols. Smaller branches are expanded from the main node, and these branches can be extended outward to new branches, until finally reaching every digital device or terminal connected to the Internet.

The communication path connecting cloud users and their service providers may include multiple ISP networks. The net structure of the Internet connects the digital terminals that are connected through a number of selectable network routes. In actual operation, it is determined which route to choose according to the current network conditions. Therefore, when using cloud services, even if a certain network or even multiple networks fails, the communication is generally guaranteed to be uninterrupted, but it may cause routing fluctuations or delays.

Some more detailed network knowledge, we will specifically introduce in Chap. 4.

1.7.2 Data Center Technology

Compared with geographically dispersed IT resources, effectively organizing many IT resources adjacent to each other to form a data center is more conducive to energy sharing, increasing the utilization rate of shared IT resources, and improving IT personnel’s efficiency. These advantages make data centers bloom all over the world. A modern data center refers to a particular type of IT infrastructure used to centrally place IT resources, including servers, databases, network and communication equipment, and software systems.

The data center contains physical and virtual IT resources. The physical layer refers to the infrastructure where computing/networking systems and equipment, hardware systems, and operating systems are placed. The virtualization layer abstracts and controls resources and is usually composed of operation and management tools on the virtualization platform. The virtualization platform outlines physical computing and network IT resources into virtualized components, which makes it easier to allocate, operate, release, monitor, and control resources.

The data center is based on standardized commercial hardware, designed with a modular architecture, and integrates multiple identical infrastructure modules and equipment, with features such as scalability, substitutability, and the ability to quickly replace hardware. Modularity and standardization are critical conditions for reducing investment and operating costs because they can achieve economies of scale in procurement, deployment, operation, and maintenance.

Common virtualization strategies and the ever-improving capacity and performance of physical devices have promoted IT resources because fewer physical components can support more complex configurations. Integrated IT resources can serve different systems and can also be shared by other cloud users.

Data centers usually have the following characteristics.

1. Automation

   The data center has a unique platform that can automate tasks such as provisioning, configuration, patching, and monitoring without manual operations.

2. Remote operation and management

   In the data center, most of the IT resources' operation and management tasks are completed through the network remote console and management system. Technicians generally do not need to enter the dedicated room where the server is placed unless they perform equipment handling, wiring, or hardware-level installation and maintenance tasks.

3. High availability

   For data center users, any form of downtime in the data center will significantly impact the continuity of their tasks. Therefore, to maintain high availability, data centers have adopted increasingly high redundancy designs; to cope with system failures, data centers usually have redundant uninterruptible power supplies, integrated wiring, and environmental control subsystems; for load balancing, then there are redundant communication links and cluster hardware.

1.7.3 Virtualization Technology

Among the many existing definitions of cloud computing, there is a definition that describes cloud computing as "accessible through the network, accessible on-demand, subscription-paid, shared by others, packaged outside of your own data center, simple Easy-to-use, virtualized IT resources." Although this definition is not comprehensive, it at least points out that virtualization technology is essential for cloud computing.

Virtualization is a virtual (rather than real) version created for certain things, such as hardware platforms, computer systems, storage devices, and network resources. Its purpose is to get rid of the

various limitations of physical resources in reality, that is, “virtualization is a logical representation of resources, and physical limitations do not restrict it.”

Although many people are interested in virtualization technology because of cloud computing, virtualization technology is not a new technology. From the virtualization of IBM’s mainframe computers to the current VMware series of desktops by EMC (Ian Xin company, acquired VMware in 2003), stand-alone virtualization technology has experienced more than half a century of development. In the early days, virtualization technology was implemented to make a single computer look like multiple computers or completely different computers, thereby improving resource utilization and reducing IT costs. With the development of virtualization technology, the scope of the concept of virtualization is also increasing.

Computer systems are usually divided into several levels, from bottom to top, including the underlying hardware resources, operating systems, user software, etc. The emergence and development of virtualization technology enable people to abstract various underlying resources to form different “virtual layers” and provide upward with the same or similar functions as the real “layers,” thereby shielding the differences in equipment and making the underlying equipment Transparent to upper-level applications. Virtualization technology reduces the degree of coupling between resource users and resource entities so that users no longer depend on specific types of certain types of resources.

The virtualization involved in cloud computing is a higher level of virtualization after development. It means that all resources-computing, storage, applications, and network equipment are connected and managed, and scheduled by the cloud platform. With the help of virtualization technology, the cloud platform can uniformly manage the diverse resources at the bottom. It can also conveniently manage resource scheduling at any time and realize the on-demand allocation of resources so that a large number of physically distributed computing resources can be logically controlled. It is presented in an overall form and supports various application requirements. Therefore, the development of virtualization technology is a crucial driving force for cloud platforms.

Although virtualization is a crucial component of cloud computing, cloud computing is not limited to virtualization. Cloud computing also expresses the service model of on-demand supply and billing and technical characteristics such as flexibility, transparency, and building blocks. Chapter 3 of this book will describe in detail the related technologies of virtualization.

1.7.4 Web Technology

Cloud computing has a deep-rooted dependence on the Internet and Web technologies. Web technology is often used as the realization medium and management interface of cloud services.

1. Basic web technology

The World Wide Web is a system of interconnected IT resources accessed through the Internet. Its two fundamental components are the Web browser client and the Web server. Other components, such as proxies, caching services, gateways, and load balancing, are used to improve Web application features such as scalability and security. These additional components are located in the hierarchy between the client and the server.

Web technology architecture consists of three basic elements.

- Uniform Resource Locator (URL): A standard syntax used to point to Web resources' identifier. URLs usually consist of logical network locations. If you want to locate Huawei's official website, enter the URL in the browser, and you can see the homepage of the official website.
- HyperText Transfer Protocol (HTTP): The basic communication protocol for exchanging content and data through the World Wide Web. Usually, the URL is transmitted via HTTP. When a user visits a webpage, the browser will send a webpage request to the website corresponding to the URL , using HTTP, and the website will return the requested webpage to the browser after receiving the request.
- Markup Language-Markup Language: It provides a lightweight method to represent Web-centric data and Metadata. At present, Hyper Text Markup Language (HTML) is commonly used in webpages, and the meaning of its tags is fixed. For example, the tag <p > means segmentation in the page; while the user defines Extensible Markup Language (XML) tags, the user can freely assign meaning to the Web data through the metadata. Metadata refers to data describing other data, or structural data used to provide information about a certain resource. If "<author>Lu Xun< /author>"is defined in the XML file, the custom tag <author> here belongs to metadata. HTML and XML are the two main markup languages that are currently widely used.

Web resources are also called hypermedia to distinguish them from hypertext. This also means that all kinds of media, such as images, audio, video, and plain text, can be referenced in a single file. However, some types of hypermedia require additional software or Web browser plug-ins to play or watch. The Web browser can request to perform read, write, update, or delete operations on Web resources on the Internet, and identify and locate them through the URL of the resource. A request for a webpage is sent to a resource host identified by a URL through HTTP, and then the Web server locates the resource and processes the requested operation, and sends the processing result back to the browser client. Processing results generally consist of HTML or XML statements.

2. Web application

Distributed applications based on Web technology (usually displaying the user interface through a Web browser) are generally considered Web applications. Due to their high accessibility, these applications appear in all types of cloud-based environments.

A typical Web application may have a three-tier model, as shown in Fig. 1.15. The first layer is the Presentation Layer, which is used to represent the user interface. The second layer is the Application Layer, which is used to implement application logic. The third layer is the Data Layer, which consists of persistent data storage. This pattern is also commonly referred to as the Model-View-Controller (MVC) pattern.

The MVC pattern model is the part of the application used to process the data logic of the application. Usually, model objects are responsible for accessing data in the database. This is corresponding to the data layer. The core device is a Data Storage Server or a Database Server. The view is the part of the application that handles the display of data. Usually, views are created based on model data. This corresponds to the presentation layer. The browser on the client-side is used for requesting and displaying Web data; the Web server on the server-side is used to process the browser' s request and return the requested webpage (corresponding to a static website) or corresponding to the results of program execution (corresponding to dynamic webpages). The controller is the part of the application that handles user interaction. Usually, the controller is responsible for reading data from the view, controlling user input, and sending data to the model. Corresponding to the application layer here, it mainly deals with transaction logic, and the core device is the Application Server.

The MVC model helps manage complex applications, simplifies group development, and makes application testing easier.

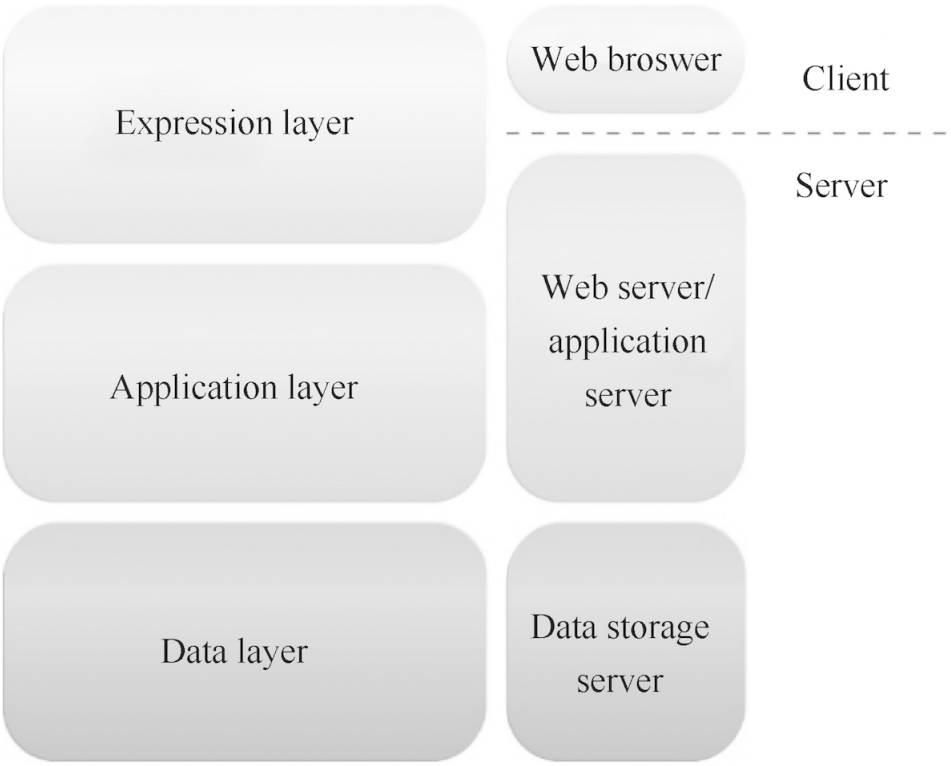

Fig. 1.15 The three-layer model of Web applications

1.7.5 Multi-Tenant Technology

- The purpose of designing multi-tenant applications is to make it possible for multiple users (tenants) to logically access the same application simultaneously. Each user has his view of the application that he uses, manages, and customizes, corresponding to a specific instance of the application. At the same time, each tenant will not realize that other tenants are using the application.

Multi-tenant applications ensure that each tenant will not access data and configuration information that is not their own. Moreover, each tenant can independently customize its application features.

- User Interface: Tenants can define application interfaces with special interface appearances.
- Business Process: When implementing applications, tenants can customize the rules, logic, and workflow of business processing.
- Data Model: Tenants can extend the application data model to include, exclude or rename the fields of the application data structure.
- Access Control: Tenants can independently control the access rights of users or groups.

Multi-tenant application architecture is usually much more complicated than single-tenant application architecture. It needs to support multi-user sharing of various components (including entrance, data model, middleware, and database), and it also needs to maintain a security level to isolate different tenants' operating environments.

The general characteristics of multi-tenant applications are as follows:

- Usage Isolation: The usage behavior of a tenant will not affect the availability and performance of the application to other tenants.
- Data Security: Tenants cannot visit data of other tenants.
- Recovery: Data backup and recovery processes are operated separately for each tenant.
- Application Upgrade: The upgrade of shared software will not have negative influence on tenants.
- Scalability: The application can be expanded according to the increase in the use demand of existing tenants or the increase in the number of tenants.
- Metered Usage: Charge based on the application processing and functions actually used by the tenant.
- Data Tier Isolation: Tenants have independent databases, tables, and schemas that are isolated from other tenants. Or, it can also be specially designed for multi-tenant shared databases, tables, and schemas.

1.7.6 Service Technology

Service technology is the foundation of cloud computing. It laid the foundation of “as a service” cloud delivery model. The typical realization and construction of service technology of cloud computing are as follows:

1. Web service

   The technical standard of Web service are as follows:

   - Web Service Description Language，WSDL: This markup language is used to create WSDL definitions, which define the application programming interface (API) of Web services, including its independent operations (functions) and input/output messages for each operation.
   - XML Schema Definition Language，XML: The messages exchanged by Web services are generally expressed in XML. This language is used to describe the XML schema. The XML schema defines the data structure of XML-based input/output messages, which are exchanged by Web services. The XML schema can be directly linked to the WSDL definition or embedded in the WSDL definition.
   - Simple Object Access Protocol，SOAP: This protocol defines the general message format of the request and response messages exchanged by Web services. A SOAP message consists of a body and a header. The body is the content of the message. The header generally contains metadata that can be processed at runtime.

- Universal Description，Discovery, and Integration，UDDI protocol: The protocol stipulates that the server must register and publish the WSDL definition to the service catalog so that users can discover the service.

The above four technologies form a classic Web service technology, and the relationship is shown in Fig. 1.16.

2. REST service

Compared with complex technologies such as SOAP and WSDL, REST (Representational State Transfer translated as representational state transfer) is a lightweight and concise software architecture style that can reduce the complexity of development and improve the scalability of the system. Mainstream cloud service providers are increasingly adopting REST-style design and implementation to provide Web services. For example, Amazon offers REST-style Web services for book searches; Yahoo! provides REST-style Web services.

REST services do not have an independent technical interface. Instead, they share a common technical interface called a uniform contract, which corresponds to a set of architectural constraints and principles. An application or design that meets these constraints and regulations is called REST. REST is usually based on existing widely popular protocols and standards such as HTTP, Uniform Resource Identifier (URI), XML, and HTML.

3. Service agent

Cloud-based environments rely heavily on service agents to perform most tasks such as monitoring and metering. These service agents are usually customized to accomplish specific tasks such as elastic expansion and pay-per-use. The service agent is an event-driven program, which intercepts messages and performs related processing at runtime.

Service agents are divided into active service agents and passive service agents. Both service agents are common in cloud-based environments. After the active service agent intercepts and reads the message, it may take certain measures, such as modifying the message content or message path. The passive service agent does not modify the content of the message, but after reading the message, it captures the specific content for monitoring, recording, or reporting.

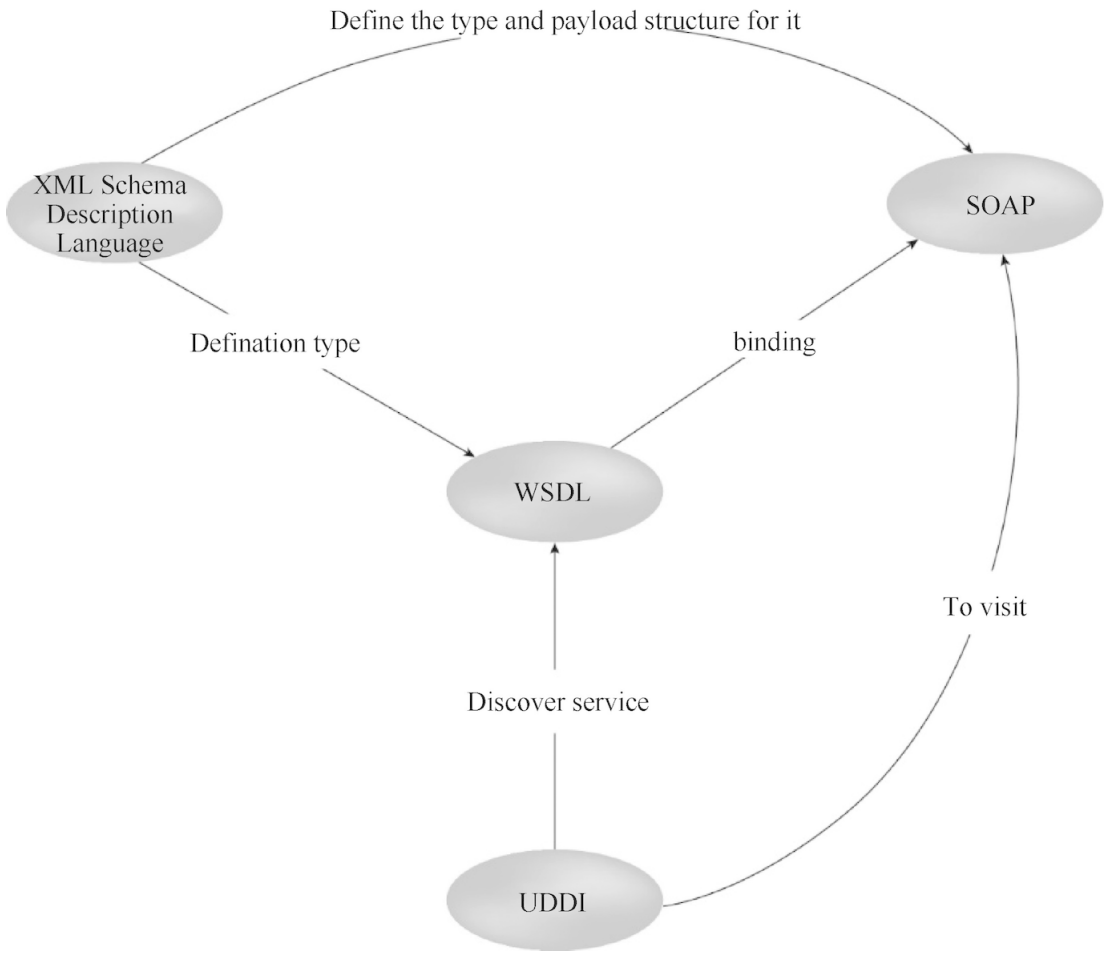

Fig. 1.16 The relationship of four Web service technologies

1.8 Understand Cloud Computing

We can understand cloud computing from multiple perspectives: cloud computing is a business model and a computing model or an implementation method. At the same time, open source technology has also been widely used in cloud computing. Open source technology and software have promoted the development of cloud computing technology, and the research on cloud computing technology has further boosted the development of open source technology. The two complement each other.

1.8.1 The Ternary Epistemology of Cloud Computing

With the development of cloud computing ecosystem, today's cloud computing should include three aspects: business model, computing model, and implementation method.

1. Cloud computing as a business model

   Cloud computing services represent a new business model. SaaS, PaaS, and IaaS are the three manifestations of this business model. For any business model and being feasible in theory, it is also necessary to ensure that it is feasible in practice. Therefore, along with the development of cloud

computing service concepts, cloud computing has also formed a set of software architecture and technical implementation mechanisms, and the cloud platform we often hear is a concrete manifestation of this mechanism.

Amazon sells all “commodities” suitable for e-commerce, including books, DVDs, computers, software, video games, electronic products, clothes, furniture, computing resources, and so on. When launching EC2, Amazon also faced many questions about “why does this retailer want to do this,” but the company’s CEO Bezos had a much broader understanding of business concepts at the time. Bezos believes that whether it is “personal computer + software” or this way of obtaining services from the “cloud,” it is not only a technical issue, but also a “business model.”

To enable the website to support large-scale business, Amazon has made excellent infrastructure construction efforts and has naturally accumulated much experience. In order to sell a large number of idle computing resources as commodities, Amazon has successively launched storage and computing rental services such as S3 (Simple Storage Service) and EC2. Bezos said, “We think this will be an exciting business someday, so our purpose of doing this is straightforward: we think this is a good business.” Although the media thinks this is a safe time for Bezos. A risky bet after the dot-com bubble, “The CEO of Amazon wants to use the technology behind his website to run your business, but Wall Street only wants him to be optimistic about his storefront.” But EC2 did affect the entire industry, and it also affected many people, the industry was obviously shaken at the time.

Before Amazon, although many services have the characteristics of cloud computing services, even the services provided by Google can still be regarded as a business model within the meaning of Internet services. After Amazon launched IaaS, it seems to have opened a window to the Internet world, telling people that computing resources can also be operated in this way, and there is a new business model called cloud computing. And those service models that are similar to traditional Internet services can finally be independent and find their own position-cloud computing services.

The University of California, Berkeley pointed out in a report on cloud computing that cloud computing refers to applications provided in the form of services on the Internet and the hardware and software that provide these services in data centers. The hardware and software are called “clouds.”

The National Institute of Standards and Technology issued Cloud Computing Synopsis Recommendations in 2011, in which PaaS, SaaS, and IaaS were explained in detail. Many people think that SaaS must run on PaaS, PaaS must run on IaaS, but in fact, there is no absolute hierarchical relationship between the three. They are all a kind of service, which can have a hierarchical stacking relationship or not.

2. Cloud computing as a computing model

From the perspective of computing mode, the earliest origin of cloud computing should be ultra-large-scale distributed computing. For example, Yahoo!‘s ultra-large-scale distributed system designed to solve system support for large-scale applications is to decompose large problems and solve them together by a large number of computers distributed in different physical locations. However, with the continuous development and improvement of technology, cloud computing draws on many other technologies and ideas, including virtualization technology and SOA concepts, when solving specific problems. Cloud computing is fundamentally different from these technologies, not only in commercial applications but also in implementation details.

As a computing model, cloud computing has its computing boundary determined by upper-level economic and lower level technical factors. Economic factors determine the business form of this computing model from top to bottom, and technical factors determine the technical form of this computing model from bottom to top.

The computing model as a cloud computing service can be further understood from two perspectives: the horizontal cloud body’s logical structure and the logical structure of the vertical cloud stack.

(a) Logical structure of horizontal cloud

The logical structure of the horizontal cloud body of cloud computing is shown in Fig. 1.17. From the perspective of the horizontal cloud, cloud computing is divided into two parts: Cloud Runtime Environment and Cloud Application.

The cloud runtime environment includes Processing, Communication, and Storage, which together support all aspects of upper cloud applications.

From this perspective, we can see that the structure of cloud computing is very similar to the structure of the personal computer we usually use. And Chaps. 3 to 5 of the book cover the 3 dimensions of processing (virtualization technology), communication (networking) and storage (distributed storage) respectively.

(b) Logical structure of vertical cloud stack

The logical structure of the vertical cloud stack is similar to the previous business model, and it is also composed of three parts: SaaS, PaaS, and IaaS, except that it will be viewed from a technical point of view.

SaaS, PaaS, and IaaS have become the “recognition cards” for people to perceive cloud computing. Many people will view the relationship between these three technical layers in a hierarchical manner. For example, SaaS runs on PaaS, and PaaS runs on Above IaaS. It can be

further seen that the IaaS layer includes Physical Hardware and Virtual Hardware; the PaaS layer includes operating systems and middleware; and there are business processes on top of the application software of the SaaS layer. The logical structure of the vertical cloud stack of cloud computing is shown in Fig. 1.18.

From a technical point of view, there is no obvious difference between SaaS users and the users of ordinary stand-alone software. PaaS provides platform services, so users are developers and need to understand the development and deployment of applications in the platform's environment. And IaaS delivers the lowest level of infrastructure services, so the users it faces are IT managers, that is, IT managers will configure and manage them first, and then perform software deployment and other tasks on it.

Although people are accustomed to dividing services according to the content provided by service providers, there is no absolute clear boundary between these three service models. Some more powerful cloud computing service providers may provide products with both SaaS and PaaS features, and some cloud computing service providers try to provide a complete set of cloud computing services, further blurring the differences in the three service models at levels.

People are slowly realizing that there are infinite possibilities of services provided through the Internet, and many companies have discovered new directions for Internet services. Therefore, in addition to SaaS, PaaS, and IaaS, some new service form names have appeared, such as Business Process as a Service, Database as a Service, and Security as a Service.

It is undeniable that these emerging cloud computing services extend the concept of Internet services and provide information services that are more in line with the laws of commercial development. If the emergence of the Internet has greatly satisfied people's needs for rapid acquisition and sharing of knowledge, then cloud computing services have met people's needs for convenient acquisition, sharing, and innovation of knowledge to a greater extent based on traditional Internet services.

After this "business model" concept of "simpler, more convenient, and lower cost" through the Internet has been widely used to meet various needs, service providers are gradually trying all services that can be provided to users through the Internet. "Cloud," so now there is a term XaaS. Where X refers to Anything or Everything, which stands for "everything can be a service." It now appears that the commercial practice of various new possibilities has continued to develop and enrich the possible meaning of cloud computing services.

(c) Cloud computing as a realization model

The ultimate realization of cloud computing requires a new generation of software/hardware technology to promote, that is, the current popular data center, and evolve toward a software-defined data center (SDDC). The data center is the ultimate home of cloud computing, including a full range of computing, storage, and communication requirements. With the data center's operation, everyone began to encounter a series of common problems, including hardware resource utilization, scalability, and automated management. Hardware upgrades take years and months, and it is usually difficult to meet the needs of a fast-developing business. Software definition is a realistic and feasible way out. Therefore, the software-defined data center has quickly become a hot keyword in the IT industry.

The software-defined data center is a relatively new concept that extends virtualization concepts (such as abstraction, centralization, and automation) to all data center resources and services to achieve IT as a Service (ITaaS). In a software-defined data center, all infrastructure elements (network, storage, CPU, and security) are virtualized and delivered as services.

The core resources of a software-defined data center are computing, storage, and networking. These three are undoubtedly the basic functional modules. Unlike traditional concepts, the software-defined data center emphasizes the capabilities abstracted from the hardware rather than the hardware itself.

For computing, computing capabilities need to be abstracted from the hardware platform, so that computing resources can break away from the hardware constraints and form a resource pool. Computing resources also need to be able to migrate within the software-defined data center's scope to adjust the load dynamically. Although virtualization is not a necessary condition, it is non-virtualization that can meet these requirements. The requirement for storage and network is the separation of the control plane and data plane. This is the first step to break away from hardware control, and it is also the initial stage of being able to define the behavior of these devices with software. After that, it is possible to consider connecting the control layer and the data layer to the software-defined data center. Security has increasingly become a factor that needs to be considered separately in data centers. Security hazards may appear between basic computing, storage, and the network or hidden in the data center's management system or the user's software. Therefore, it is necessary to regard safety as a basic functional module alone, parallel with the above three basic available modules.

Having these basic functional modules is not enough. A centralized management platform is needed to link them together, as shown in Fig. 1.19.

Automated management is the key to organizing the basic functional modules of the software-defined data center. Here it must emphasize “automated” management, not just a set of exquisite interfaces. An important driving force of the software-defined data center is the user’s management of ultra-large-scale data centers, and “automation” is undoubtedly a must.

In summary, there is no inevitable relationship between cloud computing services, cloud computing models, and cloud computing implementation. If a service implemented with traditional underlying architecture or similar to supercomputing has the three characteristics of cloud computing services: a large user base, always online, and accessibility anytime, anywhere, it can also be called cloud computing. The architecture and specific implementation itself are designed to put forward various solutions to the problems of “big users,” “big data,” and “big systems,” which are also typical problems encountered when providing cloud computing services. Therefore, cloud computing services supported by cloud computing architecture and implementation can improve service efficiency and give full play to the capabilities and advantages of cloud computing.

Just like the evolution of species, society itself will continue to advance and develop, and as a result, different progressive service models and technical needs will be generated. People’s demand for computing has promoted the popularization and development of computers, and the demand for communication and sharing has promoted the birth of the Internet. Cloud computing is also a result of social demand. With the desire for knowledge, continuous innovation and sharing, people continue to put forward new information services and products. The emergence of cloud computing, on the one hand, solves the increasingly prominent pressure problem at the system level; on the other hand, it broadens the scope of network applications and the possibility of innovation and further satisfies the premise of significantly reducing the cost of people creating and sharing knowledge. The needs of human society to acquire, innovate, and share knowledge. Therefore, cloud computing is an inevitable product of the development of the information society. With the development of the application environment, cloud computing will become more and more popular, which will bring a new experience of information society to humanity.

One of the significances of the “industrial revolution” is to free people from the shackles of production conditions and greatly liberate the productivity of material products and tangible services. The emergence of cloud computing is also gradually freeing people from the constraints of using computing resources and information services, reducing the cost of knowledge acquisition, making knowledge generation easier and sharing more convenient. It revolutionized

the productivity of information products and knowledge services. Therefore, cloud computing is as important as steam engines, internal combustion engines, and electricity and will bring about an industrial revolution in the information society.

Nowadays, cloud computing is still developing, and the extent to which it will develop in the future is still unknown. We are still exploring and deepening our understanding of cloud computing. After all, we have a process of "hearing and knowing" for new things. For cloud computing at this stage, what is most needed is support, and what is most feared is to belittle, or to be contemptuous of conclusions. But in any case, cloud computing has already had a positive impact on some areas of human society's production and life. I believe that with the development of technology and service innovation, the "cloud computing era" will come soon, and ultimately affect each of us.

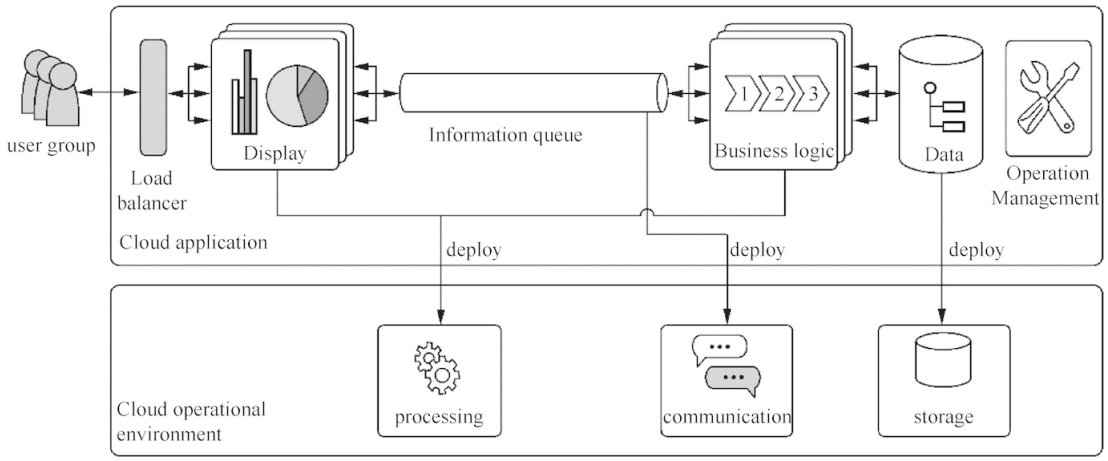

Fig. 1.17 Cloud computing's logical structure of horizontal cloud body

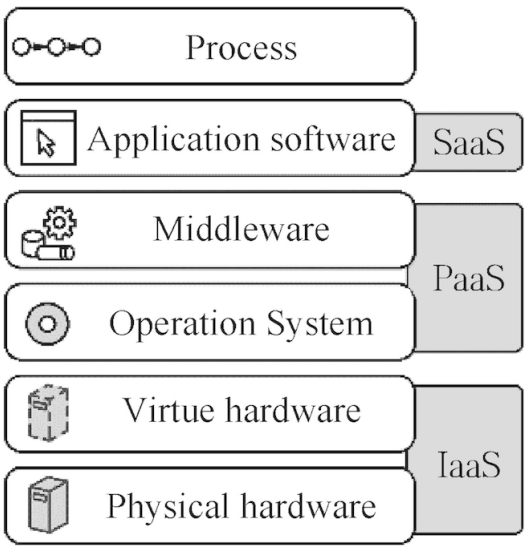

Fig. 1.18 Logical structure of vertical cloud stack for cloud computing

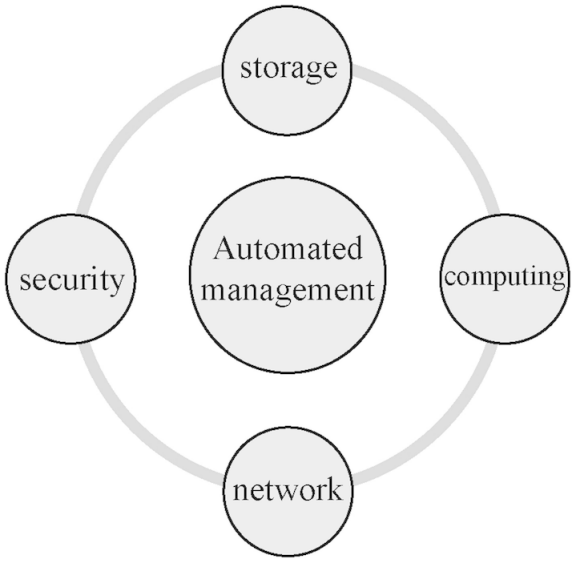

Fig. 1.19 Software-defined data center function division

1.8.2 Open Source Methodology of Cloud Computing

Open source technology has been widely used in the field of cloud computing. In the era of cloud computing, open source is not only an open source software product, but also a methodology and a collaborative way to construct large-scale and complex software.

1. The definition of open source and relative concepts

   Open source refers to the opening of the source code, source data, source assets of a type of technology or a product, which can be technologies or products of various industries, and its scope covers multiple social dimensions such as culture, industry, law, and technology. If it is software code that is open, it is generally called open source software, or open source software for short. Open source's essence is to share assets or resources (technology), expand social value, improve economic efficiency, and reduce transaction barriers and social gaps. Open source is closely related to open standards and open platforms.

   Open source software is a kind of computer software that the copyright holder provides anyone with the right to study, modify, and distribute, and publish the source code. The Open Source Initiative (OSI) has a clear definition of open source software, and it is recognized in the industry that only software that meets this definition can be called open source software. This name stems from the proposal of Eric Raymond. OSI defines the characteristics of open source software as follows:

- The license of open source software should not restrict any individual or group from selling or gifting broad-based works containing the open source software.
- The program of open source software must contain source code, and the release of source code and subsequent programs must be allowed.

- Open source software licenses must allow modification and derivative works and be published using the original software's license terms.

  An open source license is a license for computer software and other products that allows source code, blueprints, or designs to be used, modified, or shared under defined terms and conditions. At present, there are 74 kinds of OSI-certified open source licenses, and the most important are only 6–10 kinds (the most important two are GPL and Apache). Under the wave of open source commercialization, moderately loose Apache and other licenses are more popular.

  Free software is software that users can freely run, copy, distribute, learn, modify, and improve. Free software needs to have the following characteristics: no matter what purpose the user is in, he can freely run the software according to his own wishes; the user can freely learn and modify the software to help users complete their calculations as a prerequisite. The user must have access to the source code of the software; the user can freely distribute the software and its modified copies, and the user can share the improved software with the entire community for the benefit of others.

  Free software is a kind of free computer software that the developer owns the copyright and reserves the right to control the distribution, modification, and sale. The source code is usually not released to prevent users from modifying the source code.

  In a broad sense, free software is a subset of open source software, and the definition of free software is stricter than that of open source software. At the same time, open source software requires the source code to be attached when the software is released, and it is not necessarily free; similarly, free software is just the software provided to users for free, not necessarily open source. The relationship between open source software, free software, and free software is shown in Fig. 1.20.

  The open source software market is widely used. According to a Gartner survey, 99% of organizations use open source software in their IT systems. At the same time, open source software is widely used in server operating systems, cloud computing, and the Web.

  The scale of the open source software market ranks first among server operating systems. According to statistics, more than 90% of the operating systems running on global public cloud vendors' servers are open source Linux operating systems. Its market share in the embedded market is 62%, and its market share in the field of supercomputing is even higher. Reached 99%. The Android system based on the Linux kernel runs on more than 80% of smart phones in the world.

  The 2019 annual report released by GitHub, the world's largest developer community, revealed a data: GitHub currently has more than 40 million developer users worldwide, 80% of which are

from outside the United States, and the use of open source in China is proliferating. In 2018 alone, nearly 10 million new developer users joined the GitHub community and contributed to 44 million open source projects worldwide.

Open source software is also widely used in the field of cloud computing. Open source in the cloud computing field is currently mainly based on two levels: IaaS and PaaS. The IaaS level includes OpenStack, CloudStack, oVirt, ZStack, etc., and the PaaS level includes OpenShift, Rancher, Cloud Foundry, and the scheduling platform Kubernetes, Mesos, etc. For example, OpenStack software is widely used in IT, telecommunications, research, finance, and other fields. Another example is the open source application container engine Docker. Since its release in 2013, its technology has become increasingly mature. At present, the number of container image downloads has exceeded eight billion times.

The wide application of open source software in other fields also includes big data, Software Defined Network (SDN), Network Function Virtualization (NFV), artificial intelligence, and other areas. For example, big data basic analysis platforms include Hadoop, Spark, etc., NFV has OPNFV, and artificial intelligence has TensorFlow.

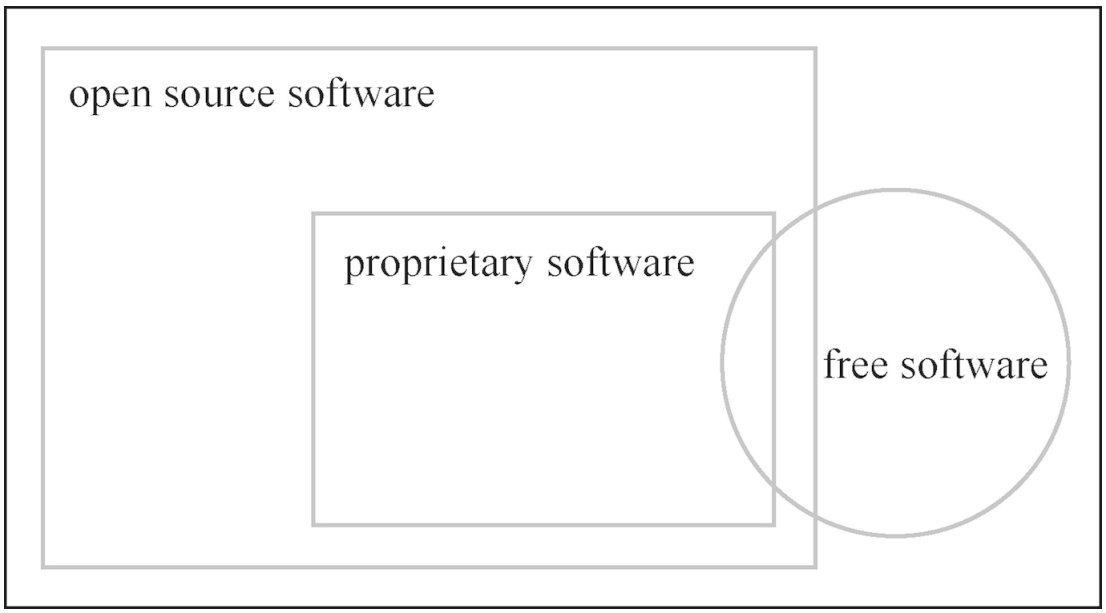

Fig. 1.20 The relationship between open source software, proprietary software and free software

1.8.2.1 The Significance of Open Source

1. The open source ecology promotes national information technology innovation and drives economic development.

Open source effectively promotes technological innovation. The open source model can effectively achieve information exchange, obtain newer source codes of key technologies, use global

technical resources to promote technological development iterations rapidly, break technical barriers, and promote new technologies' popularization.

Open source can realize software autonomous and controllable. Open source is more transparent and open. The establishment of my country's open source software industry can effectively guarantee information security, achieve autonomy and control, and ensure that information security is easier to manage. Products and services generally do not have malicious backdoors, and vulnerabilities can be continuously improved or patched.

Open source promotes the development of education and scientific research. Open source provides more independent learning resources for teachers and students in colleges and universities. Students can join open source projects directly and quickly, with continuous technical level improvement and continuous enrichment of experience.

Open source promotes the development of industry informatization. The open source model can effectively reduce application costs and technical thresholds, accelerate enterprise informatization, and promote the vigorous development of my country's economy.

2. Software vendors rely on open source technology to enhance R&D capabilities.

Software vendors use open source technology to reduce research and development costs, communicate through open source communities, and are familiar with the use of open source technology, which facilitates tracking of open source technology updates and adapts them according to business needs.

Well-known open source projects have the participation of high-level R&D personnel in the industry, and their source code has many references for technical personnel in terms of coding style and algorithm ideas. In using open source projects or conducting secondary development based on open source projects, software manufacturer R&D personnel can learn innovative methods to solve problems by reading source code and other methods.

After software vendors open the project, the project has a wider range of users and more complex application scenarios. R&D personnel should consider the company's business needs and personnel usage when developing but need to pay more attention to code compatibility and standardization.

3. Users use open source technology to change the route of informatization.

Enterprise users can carry out customized development based on open source technology. The functions that end-user information systems need to achieve are different. Compared with closed-source software, open source software is more flexible and has a higher degree of customization. End enterprise users can do secondary development based on open source code to achieve the requirements of specific scenarios and specific functions and avoid binding risks.

Use open source technology to allow companies to focus on innovation. As more business resources get rid of developing software shackles, the focus of enterprises will shift to innovation. Creativity flourishes in small- and medium-sized enterprises because they are better able to create competitive alternative technologies and proprietary software to obtain more unique and forward-looking thinking than their competitors.

4. Enterprises independently open source and lead the technological development path.

Enterprises' independent open source can effectively improve R&D efficiency and enhance the quality of code. The project's open source process can attract outstanding developers and users to participate in it, inject more "fresh blood," and allow the project to continue to develop. At the same time, open source projects are deployed in different application scenarios, exposing more problems in the project and saving test costs.

Enterprises' independent open source can lead the development of technology and establish an ecosystem with open source enterprises as the core. During the operation of open source projects, potential users can be attracted to use open source software, so that more companies and developers in the industry can understand the technological development of the companies that belong to the open source project, and establish an upstream and downstream ecosystem of providers and users through open source technology, and keep abreast of users. Demand, seize the business territory, and drive the healthy development of the enterprise.

1.8.2.2 Open Source Is a Methodology

The vital impact of open source is that it makes learning programming easier. Any novice can access countless mature products for free as a reference, and the novice will one day become an experienced developer and feedback the open source community. Therefore, the open source community can develop continuously and sustainably, and the open source culture has become the programmer community's representative culture.

Open source has two aspects: one is open source software technology and related aspects, including open source software history, open source software agreements, technical products, open source communities, related hardware, technical personnel, open source software-related industries and enterprises. The second is open source values and methodology. Relevant content includes open source value system, open source methodology system, non-technical projects carried out with open source methodology, related non-technical organizations, communities, and people. Open source values' connotation mainly includes six aspects: dedication, sense of gratitude, open spirit, courage, pursuit of continuous progress, and the spirit of obtaining fair value returns based on labor. The connotation of open source methodology mainly includes promoting progress and innovation through open sharing,

solving complex and systemic problems through gathering and accumulating the labor and wisdom of many references, completing open source projects through community platforms, and through effective organization and organization of well-known companies and individuals to develop and complete the project. Open source values and open source methodology are valuable spiritual wealth contributed by open source technology to mankind.

1.8.2.3 Open Source Brings Challenges to Cloud Computing

The open source development model provides a new way for the revolution and transformation of the industry model, and the open source software resources provide directly usable software technologies, tools, and products for the development of the IT industry. The open source development model's openness and transparency help to quickly gather public wisdom and effectively promote the formation and development of technology and application ecology.

Since the development of open source, the types of software products have become more and more diverse, their functions have become more and more powerful, and the scale of the community has become larger and larger. Accordingly, the tools and technical systems supporting open source development have become more difficult to master. The new situation has brought challenges to the further prosperity and development of open source software and put forward higher requirements for open source participants. In terms of participation in open source software development, studies have pointed out that the number and proportion of long-term contributors in well-known open source communities has declined in recent years. The survey found that in the application of open source products and technologies, quite a few IT companies lack talents who master open source technologies. Only by effectively cultivating open source talents can the team of open source contributors be expanded.

Compared with the traditional software development field, open source has different or even broader requirements for software talents, so that the conventional talent training system may need to adapt to the needs and pay attention to specific aspects of open source talent education.

1.9 Exercise

(1) Multiple Choice

1. The "cloud" in "cloud computing" is more credible about its origins ().

   A. Some aspects of cloud computing are as elusive as the cloud.
   B. The supporting technology of cloud computing is often represented by a cloud-like pattern on the Internet, so the network that provides resources is often called "cloud."
   C. The scale of cloud computing is generally as broad as the cloud.

D. Users cannot see the resources contained in cloud computing, as if hidden behind the cloud.

2. () is not a major feature of cloud computing.

   A. On-demand self-service.
   B. Extensive network access.
   C. Complementary resources.
   D. Fast elastic scaling.

3. Systems or services that do not rely on cloud computing are ().

   A. Baidu net disk.
   B. Attention to the conference system.
   C. There are Cloud Notes.
   D. The remote login system for the supercomputer.

4. The misconception about resource pooling is ()

   A. Resource pooling is one of the prerequisites for on-demand self-service.
   B. Resource pooling is equivalent to resource classification.
   C. Resource pooling requires that all resources are decomposed to a minimum unit.
   D. Resource pooling masks the differences between different resources.

5. The misconception about the rapid elastic scaling of cloud computing is () .

   A. Elastic expansion is considered one of the core reasons to engage users in cloud computing.
   B. Rapid elastic scaling means that cloud users can automatically and transparently scale their IT resources according to their needs.
   C. Rapid elastic scaling must be manually expanded or reduced.
   D. Rapid elastic scaling enables users to save money while keeping their business or applications running smoothly.

6. The misconception about the metering service for cloud computing is ().

   A. Metering is the basis of billing.
   B. Services in cloud computing are measured based on the time of use.
   C. Billing management systems are commonly available in cloud computing systems and are designed to collect and process usage data.
   D. Using a quota billing system prevents further usage requests from cloud users when the quota is exceeded.

7. () is not a service provided by cloud computing.

   A. IaaS.
   B. PaaS.
   C. SaaS.
   D. RaaS.

8. () is not a key driver of the birth and development of cloud computing technology.

   A. Increase in network bandwidth B. The emergence of deep learning techniques.
   B. The emergence of virtualization technology D. The development of the mobile Internet.
   C. Enter the era of big data.

9. The benefits of cloud computing do not include:

   A. Cost savings.
   B. Data is instantly accessible from anywhere.
   C. Improve adaptability and scale IT needs flexibly. D. Enhance the security and confidentiality of user data.

10. Multi-tenant technology is an important support technology for cloud computing. () is not a general feature of multi-tenant applications.

    A. Use isolation.
    B. B. Data security.
    C. C. Recoverability.
    D. Scalability E. Synergy.

(2) Fill in the Blanks

1. Cloud computing technology provides computing resources, _____, and other various resources to resource users in the form of services through the network.
2. _____technology is the basic support of cloud computing. The cloud is inseparable from the _____network.
3. The English abbreviation of the Internet service provider that provides network access services for cloud services is _____.
4. According to the classification of cloud computing operation mode, cloud can be divided into _____, _____, community cloud, hybrid cloud, and industry cloud.
5.

_____provides highly scalable and on-demand IT capabilities based on hardware resources such as servers and storage in the form of services. Usually charged according to the cost of the resources consumed.

6. _____is located in the middle of the three-tier service of cloud computing, usually also called “cloud operating system,” which provides end-users with an Internet-based application development environment, including application programming interfaces and operating platforms, etc.

(3) Answer the Following Questions

1. What is the definition of cloud computing?
2. What are the features of cloud computing? What are the benefits of using cloud computing?
3. What are the types of cloud computing that can be divided into operational models? What kind of security and privacy guarantees?
4. What do cloud computing have in common with traditional host computing?
5. What are the key technologies included in cloud computing technology, called cloud enabling technology?
6. What are three service models for cloud computing? Which one does Amazon’s AWS and Microsoft’s Windows Azure belong to?
7. How to understand cloud computing from the perspective of triadic epistemology?
8. What are the differences and connections between open source software, free software, and free software?