Chapter 7
Cloud Computing

Abstract
Cloud computing is an evolution of information technology and a dominant business model for delivering IT resources. With cloud computing, individuals and organizations can gain on-demand network access to a shared pool of managed and scalable IT resources, such as servers, storage, and applications. Recently, academics as well as practitioners have paid a great deal of attention to cloud computing. We rely heavily on cloud services in our daily lives, e.g., for storing data, writing documents, managing businesses, and playing games online. Cloud computing also provides the infrastructure that has powered key digital trends such as mobile computing, the Internet of Things, big data, and artificial intelligence, thereby accelerating industry dynamics, disrupting existing business models, and fueling the digital transformation. Still, cloud computing not only provides a vast number of benefits and opportunities; it also comes with several challenges and concerns, e.g., regarding protecting customers' data.

Learning Objectives of this Chapter
This chapter provides a brief introduction to the very diverse domain of cloud computing. Besides understanding how cloud computing emerged from technology improvements and innovations in recent decades, students should get to know essential characteristics of cloud computing, as well as of cloud computing service and deployment models. While other chapters in this book have already introduced several important technologies and related concepts, students will learn in this chapter how these technologies are used in cloud computing to form a multi-layered cloud stack. Finally, this chapter will provide information on the benefits and opportunities of cloud computing, while also pointing out the pitfalls and drawbacks of cloud services, such as those regarding safeguarding customers' data stored in the cloud.

Structure of this Chapter
This chapter proceeds as follows: First, a brief history of cloud computing will introduce readers to the topic. Second, the concept and building blocks of cloud computing will be described, major actors in the market will be explained, and cloud computing will be compared against related concepts. Third, essential cloud

A. Sunyaev, *Internet Computing*, https://doi.org/10.1007/978-3-030-34957-8_7

technologies and technical layers that together enable the provision of cloud services are presented. Fourth, the chapter takes an in-depth look at opportunities and challenges of cloud services for organizations, as well as at transformative mechanisms of cloud computing that enable truly innovative services and business models. The chapter explains how, ultimately, these services and models lead to fundamental and large-scale innovations that benefit individuals, organizations, markets, and societies. The chapter closes with a discussion of security and privacy challenges in cloud environments, showing how continuous cloud service certifications can help to tackle these challenges.

7.1 An Introduction to Cloud Computing

7.1.1 The Emergence of Cloud Computing

Cloud computing is a model for enabling access to computing resources that evolved in information technology and has become a dominant business model for delivering IT infrastructure, components, and applications (Benlian et al. 2018). With cloud computing, a product-centric model for IT provisioning is transformed into a global, distributed, service-centric model, leading to a disruptive shift from IT-as-a-product to IT-as-a-service (Iyer and Henderson 2010; Benlian and Hess 2011). Since its widespread emergence around 2007, cloud computing has changed the way in which IT services are invented, developed, deployed, scaled, updated, maintained, and paid for (Marston et al. 2011). Cloud computing enables individuals and organizations to access IT resources on-demand, from any device, and at any time, as a measured service (Mell and Grance 2011). It lowers the entry point to high performance computing, allowing organizations to leverage computing power for which they have neither the capital budget, nor the operational expertise to acquire (Arasaratnam 2011). While the market currently urges the availability of IT resources on-demand, cloud providers offer an ever-increasing number and variety of services that are built on a shared pool of computing resources and that can elastically scale up to growing computing demands. Market researchers predict the public cloud market will hit USD 240 billion by 2020, compared to USD 42 billion in 2010 (Gartner 2018a). Today, we heavily rely on cloud services in our daily lives, as in storing data (e.g., *Dropbox, OneDrive*), writing documents (e.g., *Office 365, Google Docs*), managing businesses (e.g., *SAP By*Design), and playing games online (e.g., *GamingAnywhere, Stadia*). Cloud computing also provides the infrastructure that has powered key digital trends including mobile computing, the Internet of Things, big data, and artificial intelligence. Thereby, cloud computing accelerates industry dynamics, disrupts existing business models, and fuels the digital transformation (Bharadwaj et al. 2013; Hess et al. 2016). Defying initial concerns, cloud computing has become a critical IT infrastructure for almost every aspect of day-to-day living, and it will continue to transform the world we live in on multiple levels and in various ways, as later sections of this chapter will show.

The history of computing makes it clear that the arrival of the cloud computing era is an evolutionary development (Iyer and Henderson 2010). Cloud computing has its roots in the advancement of several technologies, especially in hardware (e.g., virtualization, multi-core chips), Internet technologies (e.g., web services, service-oriented architectures), distributed computing (e.g., clusters, grids), and systems management (e.g., autonomic computing, data center automation) (Voorsluys et al. 2011). The idea of providing computing *as a service* through networks dates back to the late 1960s and became a driving force of the early Internet development (Venters and Whitley 2012; Berman and Hey 2004). In 1969, *Leonard Kleinrock*, who is known for the ARPANET project, which is considered to be the foundation of the Internet, said: *"As of now, computer networks are still in their infancy but as they grow up and become sophisticated, we will probably see the spread of 'computer utilities' which, like present electric and telephone utilities, will service individual homes and offices across the country"* (Buyya et al. 2009; Kleinrock 2005). This vision became reality with the emergence of 'application service provision' during the 1980s (Venters and Whitley 2012; Durkee 2010). With this service-oriented model, third-party providers deploy, manage, and remotely host packaged software applications through centrally located servers and deliver them to organizations through a rental or lease arrangement (Iyer and Henderson 2010). Organizations adopted these application services to reduce complexity, improve organizational agility, and minimize costs by paying only for what they use. However, early application service providers often failed, due to insufficient bandwidth and computing power (Susarla et al. 2003).

During the late 1990s the telecommunications industry experienced tremendous investment and entry of new organizations, which led to a large increase in global fiber-optic networking. Such changes dramatically reduced network latency and related costs (Hogendorn 2011). This improvement of networking was coupled with the emergence of technologies and techniques to coordinate the large-scale, on-demand provision of computing resources (Venters and Whitley 2012), achieved by, i.e., drawing on innovations around 'grid computing' (Foster and Kesselman 2004), 'utility computing' (Bunker and Thomson 2006), and virtualization of hardware (Killalea 2008). For example, grid computing refers to the paradigm of sharing, selecting, and aggregating large-scale geographically-distributed and (possibly) virtualized resources (software, data, computers) depending on availability, capability, cost, and related quality requirements for solving large-scale problems (Foster and Kesselman 2004; Schwiegelshohn et al. 2010). Today, grid computing is used particularly for solving resource-intensive research problems (Kroeker 2011). Section 7.1.6 will provide a detailed comparison of related concepts. Simultaneous to this development, besides advances in networking, technologies, and techniques, organizations such as *Alphabet (became the parent company of Google)*, *Amazon,* and *Microsoft* set up large data centers to transfer computing processes from individual computers and private IT infrastructures to large external and public data centers that were accessible via the Internet. Such data centers' services soon were labeled 'cloud computing' (Venters and Whitley 2012).

The evolution of distributed and service-oriented architectures to provide on-demand computing resources via the Internet, and the vast improvements of technological infrastructures, including networks and data centers, ultimately enabled a shift to the cloud (Carr 2008). The cloud provides computing by means of large pools of automated and scalable computing resources and related applications (Cafaro and Aloisio 2011; Cusumano 2010).

7.1.2 *Definition of Cloud Computing and its Essential Characteristics*

There are numerous definitions and explanations of cloud computing (Marston et al. 2011; Leimeister et al. 2010; Lins et al. 2019a). The definition put forward by the National Institute of Standards and Technology (NIST) has become established as the most commonly accepted definition of cloud computing among experts. According to this definition, cloud computing is a model that enables ubiquitous, convenient, on-demand access to a shared pool of configurable computing resources that can rapidly be provisioned at any time and from any location via the Internet or a network (Mell and Grance 2011). This includes, e.g., access to networks, servers, storage, or applications. Cloud services are rapidly deployed with minimal management effort, little interaction with the cloud provider, and they can be customized to meet the needs of cloud customers.

Cloud Computing
Cloud computing is a model that enables ubiquitous, convenient, on-demand access to a shared pool of configurable computing resources that can rapidly be provisioned at any time and from any location via the Internet or a network (Mell and Grance 2011).

Characteristic features of cloud services, as described below, include service-based provision of IT resources, on-demand self-service, ubiquitous access, multitenancy, location independence, rapid elasticity, and pay-per-use billing.

Service-Based IT-Resources
All cloud offerings can be expressed as a service. A service is based on a contract, commonly referred to as a Service Level Agreement, which defines the functions it offers and commits to upholding certain quality parameters, e.g., ones regarding service availability (Kumar and DuPree 2011).

On-Demand Self-Service
On-demand self-service enables cloud customers to independently and almost immediately provision computing capabilities, such as server time and network storage (Mell and Grance 2011). Notably, this can be done automatically and

without human interaction by the cloud providers. It is thus possible, e.g., to increase or decrease computing, storage, or application licenses customers have obtained, depending on their current needs.

Ubiquitous Access
Cloud services are provided via a broadband network, mostly using the Internet (Mell and Grance 2011). Cloud services utilize standardized communication interfaces and can be used with a variety of devices, including smartphones, tablets, and laptops. Consequently, cloud customers can access any cloud service from any platform or device at any time (Iyer and Henderson 2010).

Multitenancy
Multitenancy is the ability to have multiple customers leverage shared resources (Kumar and DuPree 2011). Thus, the resources that the cloud provider makes available are used simultaneously by multiple cloud customers (Mell and Grance 2011). Physical and virtual resources are dynamically assigned and reassigned as needed to different cloud customers through a multi-client architecture. Sharing computing resources is part of what could makes cloud computing economically beneficial (Arasaratnam 2011).

Location Independence
When using cloud services, there is a sense of location independence in that the cloud customer generally has no control over or knowledge of where the provided resources are actually located. However, on a higher level of abstraction (e.g., country, state, or datacenter), it could be possible to specify location (Mell and Grance 2011). Today, organizations with large data sets typically employ experts whose role is solely to know where data elements exist within their databases, as these organizations have to fulfill various data protection and legal requirements (Iyer and Henderson 2010).

Rapid Elasticity
Most organizations plan their IT processing capacity to cope with peak loads, which is why much of this capacity remains idle for large parts of the time (Iyer and Henderson 2010). In contrast, provisioned cloud resources can be adapted and shared more flexibly, in some cases fully automatically, in order to match the resources to the current needs (Mell and Grance 2011). Due partly to such rapid elasticity, cloud customers have the impression that resources are almost unlimited and are available at any time, to any extent. Cloud providers, therefore, need to create solutions that meet cloud customers' quality of service expectations and cope with large capacity increases, including potentially unforeseeable events (Kumar and DuPree 2011).

Pay-Per-Use Billing
With cloud services, the expectation is that cloud customers are charged for the amount of time they actually use the resource or related measures, a system referred to as 'pay-per-use' (Arasaratnam 2011). Thereby, cloud computing eliminates up-front commitment by users and changes the entry barrier for high performance

computing resources by allowing cloud customers to use only what they need for the time they need it. Cloud providers have to implement features that allow efficient service provision, such as for pricing, accounting, and billing (Buyya et al. 2009). Accordingly, metering should be done for different types of service (e.g., storage or processing) and usage needs to be reported promptly, thus providing greater transparency (Mell and Grance 2011).

7.1.3 The Cloud Service Market

The cloud service market comprises five major actors: cloud customers, cloud providers, cloud data center operators, cloud auditors, and cloud brokers. Each actor is an entity (a person or an organization) that participates in a transaction or process and performs tasks in cloud computing (Liu et al. 2011). Below, each actor is briefly described.

Cloud Customer
A cloud customer is a person or organization that maintains a business relationship with, and uses the service of a cloud provider; thus cloud customers are cloud providers' principal stakeholders (Liu et al. 2011). In business-to-business contexts, organizations want an overview of service offerings available in the market before they contact cloud providers who have potentially suitable services that could fulfil the organization's requirements. Organizations should negotiate a contractual relationship regarding the provision of the requested cloud service. Typically, such contracts take the form of service level agreements (SLA). SLAs can cover terms regarding the quality of service, security, data protection, and remedies for performance failures (Liu et al. 2011). Nevertheless, a cloud provider's pricing policy and SLA are mostly standardized and typically fixed, unless the organization expects heavy usage and thus is in a better negotiating position. In business-to-customer contexts, individuals can easily access the cloud service and download applications relevant to them by agreeing to standardized SLA. Cloud services for individuals (e.g., students) typically allow for trial versions and are offered on the basis of a freemium business model (meaning that the basic service features are free of charge, whereas payment is demanded for additional (premium) features, services, or virtual goods). The customers can be billed for the services or for premium features in a 'pay-per-use' manner, and are required to arrange payments accordingly.

Cloud Customer
A cloud customer represents a person or organization that maintains a business relationship with, and uses the service a cloud provider offers (Liu et al. 2011).

Cloud Provider
A cloud provider is a legal person or an organization that is responsible for making a cloud service available to interested parties (Liu et al. 2011). Depending on the service and deployment model, a cloud provider acquires and manages the computing infrastructure required for providing the services, runs the cloud software and operating systems, and provisions the cloud services to cloud customers through a network (e.g., the Internet). Cloud providers' activities include service deployment, service orchestration, cloud service management, service maintenance and upgrades, as well as ensuring security and privacy.

Cloud Provider
A cloud provider is a legal person or an organization that is responsible for making a cloud service available to interested parties (Liu et al. 2011).

Cloud Data Center Operators
Cloud data center operators provide the basic IT infrastructure, including server rooms, redundant or backup components, infrastructure for power supply, data communication connections, environmental controls (e.g., cooling, air conditioning, and fire emergency equipment), and various security devices. Currently, the largest data center operators include *Equinix*, *Digital Realty*, *Level 3*, *Telehouse*, *NTT*, and *Global Switch* (Cloudscene 2018). For example, there are more than 427 data centers in Germany, which are mainly located in and around Frankfurt. Data centers feature rich ecosystems and state of the art equipment, which ensures maximum uptime and connectivity to hundreds of Internet service providers, cloud service providers, and customers.

Cloud Data Center Operators
Cloud data center operators provide the basic IT infrastructure, including server rooms, redundant or backup components, infrastructure for power supply, data communications connections, environmental controls (e.g., cooling, air conditioning and fire emergency equipment), and various security devices.

Cloud Certification Authorities and Auditors
A certification authority is an independent party that assesses whether a cloud service fulfills given requirements (Liu et al. 2011). Certification authorities typically employ auditors to endorse a cloud service as compliant with the set certification criteria, such as criteria imposed by the cloud service certification standard *ISO/IEC 27018* (Sunyaev and Schneider 2013; Lins et al. 2019a). During a certification process, auditors perform comprehensive, manual checks (i.e., document reviews, on-site audits, and penetration tests) to assess adherence according to a defined set of

certification criteria (Lansing et al. 2018). If a cloud provider adheres to specified criteria, a certification authority awards a formal written certificate, which allows providers to embed a graphical web assurance seal on their websites. Cloud service certifications typically aim to ensure the availability, integrity, and confidentiality of provisioned cloud services for a validity period of one to three years (Schneider et al. 2014). Nowadays, auditors can use automated monitoring and auditing techniques, as well as mechanisms for transparently providing certification-relevant information that will continuously confirm adherence to certification criteria (Lins et al. 2018; Lins et al. 2016a).

Cloud Auditors
A cloud auditor is a party that performs an independent assessment to determine whether a cloud service fulfills given requirements (Liu et al. 2011).

Cloud Brokers
As cloud computing evolves, integrating cloud services can be too complex for cloud customers to manage (Liu et al. 2011). A cloud customer can request cloud services from a cloud broker, instead of contacting each cloud provider directly. A cloud broker is an entity that manages the use, performance, and delivery of cloud services and negotiates relationships between cloud providers and cloud customers. In general, a cloud broker can provide services in three categories (Liu et al. 2011) explained below.

- *Service Intermediation:* A cloud broker enhances a given service by improving certain specific capabilities and providing value-added services to cloud customers. For example, a cloud broker can manage access to cloud services, provide identity management, provide performance reporting, enhance security, or handle the billing process.
- *Service Aggregation:* A cloud broker combines and integrates multiple services into one or more new services (referred to as Multi-Cloud). The cloud broker provides integration capabilities that link the cloud customer and multiple cloud providers, thus enabling data integration and secure data transmission, as well as additional service intermediation capabilities.
- *Service Arbitrage:* Service arbitrage is similar to service aggregation, excepting that the services being aggregated are not fixed by any agreements with cloud customers. Service arbitrage ensures that a broker has the flexibility to choose services from multiple cloud providers. The cloud broker, for example, can constantly use a performance or security scoring service to measure and select a cloud service with the best score at a given point in time. Consider a cloud customer wanting to store 100 TB of data, who contacts a cloud broker to fulfill this demand. Based on his expertise and position in the market, the cloud broker will select the best or cheapest option to meet this demand.

Cloud Brokers
A cloud broker is an entity that manages the use, performance, and delivery of cloud services, and that negotiates relationships between cloud providers and cloud customers (Liu et al. 2011).

7.1.4 Cloud Computing Service Models

Cloud services are typically divided into three models (see Fig. 7.1), hierarchically organized according to the abstraction level of the capability provided and the provider's service model. The three models are (1) Infrastructure as a Service (IaaS), (2) Platform as a Service (PaaS), and (3) Software as a Service (SaaS) (Mell and Grance 2011). The different providers that we witness today, have particularly developed competences related to these different technical layers (i.e., software, platform, and infrastructure) (Marston et al. 2011).

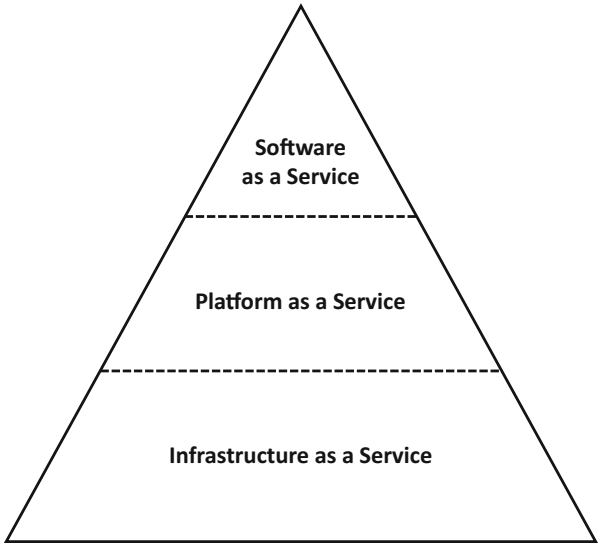

Fig. 7.1 Cloud service models depicted as a pyramid

Infrastructure as a Service (IaaS)
Cloud customers can procure processing, storage, networks, and other fundamental computing resources by using IaaS services. They can utilize these resources by deploying and running arbitrary software, which can include operating systems and applications (Mell and Grance 2011). The provider typically provides this service by dividing a very large physical infrastructure resource into smaller virtual resources for access by the customer (Arasaratnam 2011). Sometimes the cloud service the

provider offers is a complete virtual machine with an operating system. In other instances, the cloud service offers storage capacities or a bare virtual machine with no operating system. Cloud customers do not manage or control the underlying cloud infrastructure, but they have control of the operating systems, storage, and deployed applications. Also, customers' limited control of select networking components (e.g., host firewalls), is possible (Mell and Grance 2011). For example, *Amazon Web Services* mainly offers IaaS by providing virtual machines with a software stack that can be customized similarly to a conventional physical server (Voorsluys et al. 2011). Cloud customers are given privileges to perform various activities on the virtual machine, such as starting and stopping it, customizing it by installing software packages, and configuring access permissions and firewall rules.

Examples for IaaS include: backing up and recovering file systems and data stored on servers and desktop systems; delivering server resources for running cloud-based systems that can be dynamically provisioned and configured as needed; managing cloud infrastructure platforms; providing massively scalable storage capacity services that can be used for applications, backups, archival, and file storage (Liu et al. 2011). Leading cloud providers offering IaaS, among others, are *Amazon*, *IBM*, *Microsoft*, *Rackspace*, *NTT*, *Oracle*, and *Fujitsu* (ITCandor 2018).

Platform as a Service (PaaS)

Cloud customers can deploy customer-created or acquired applications to the cloud infrastructure by using programming languages, libraries, services, and tools supported by the cloud provider (Mell and Grance 2011). A cloud customer does not manage or control the underlying cloud infrastructure (e.g., the network, servers, operating systems, or storage), but can control the deployed applications and also the configuration settings for the application-hosting environment. Thereby, PaaS offers a highly integrated environment on which developers design, build, test, and deploy custom applications without the costs and complexity of buying and managing the underlying hardware and software layers (Schneider and Sunyaev 2016). *Google's App Engine* is an example of a PaaS that offers a scalable environment for developing and hosting web applications (Voorsluys et al. 2011). Its building blocks are, among others, an in-memory object cache, a mail service, an instant messaging service, an image manipulation service, and integration with Google's account authentication service.

Further examples of PaaS offerings, are platforms for application development and testing, services offering scalable relational or NoSQL databases, dedicated development platforms for building integration applications in the cloud and within the enterprise, and runtime environments suited to run specific applications (Liu et al. 2011). Leading providers offering PaaS, among others, are *Amazon*, *Microsoft*, *Alibaba*, *Google*, *IBM*, and *Rackspace* (Gartner 2018b).

Software as a Service (SaaS)

In this model, cloud customers use a cloud provider's applications running on a cloud infrastructure (Mell and Grance 2011). Applications are typically accessible from various client devices through either a thin client interface, such as a web browser, or a program interface. The customers do not manage or control the

underlying cloud infrastructure (e.g., the network, servers, operating systems, storage, or even individual application capabilities), with the possible exception of limited user-specific application configuration settings. Thus, a cloud customer, in most cases, is completely abstracted from the nuances of the application running behind the scenes (Arasaratnam 2011). Due to SaaS, customers are increasingly shifting from locally installed computer programs to online software services that offer the same functionality while also alleviating customers' possible burden of software maintenance (Voorsluys et al. 2011).

Examples of SaaS are: applications for email, word processing, spreadsheets, and presentations; application services to manage customer billing based on usage and subscriptions to products and services; customer relationship management applications that range from call center applications to sales force automation; tools that allow users to collaborate in workgroups within enterprises and across enterprises; applications for managing financial processes ranging from expense processing and invoicing to tax management. Leading cloud providers offering SaaS, among others, are *Salesforce*, *Microsoft*, *SAP*, *Oracle*, *Adobe Systems*, and *IBM* (Apps Run The World 2017). The company Loudcloud is supposed to be one of the first vendors that talked about cloud computing and SaaS back in 1999.

Besides the fundamental classification of cloud service models into PaaS, IaaS, and SaaS, there are many other service models in practice and the literature. The latter are often summarized as Everything as a Service (XaaS) (Singh et al. 2016). Table 7.1 lists examples of such service models, including Database as a Service or Security as a Service, and shows how they relate to the fundamental service models (see also Höfer and Karagiannis (2011)).

Table 7.1 Cloud service models and their assignment to the fundamental service models, software, platform and infrastructure as services

Service model	Fundamental service models			Exemplary Literature
	SaaS	PaaS	IaaS	
Security as a service	●	-	-	Sharma et al. (2016)
Search as a service	●	-	-	Dašić et al. (2016)
Testing as a service	-	●	-	Linthicum (2009)
Database as a service	-	●	-	Linthicum (2009)
Network as a service	-	-	●	Soares et al. (2011)
Rendering as a service	-	-	●	Annette et al. (2015)

7.1.5 Cloud Computing Deployment Models

The NIST definition of cloud computing distinguishes between four basic deployment models: private, public, community, and hybrid cloud (Lins et al. 2019a; Mell and Grance 2011). In addition, virtual private and multi-cloud deployment models are often discussed in the literature and practice.

Private Cloud Models
The private cloud infrastructure is used only by a single person or organization and its members (Mell and Grance 2011). It can be owned, managed, and operated by the person or organization, by a third party, or by a combination of both. The private cloud generally serves internal company purposes, and cloud customers have full control over who, how, and when a cloud service can be used. For example, a private cloud service could be utilized by a financial company that seeks to store sensitive data, yet still wants to benefit from the advantages of cloud computing, such as on-demand resource allocation. Multiple cloud providers – including *Amazon*, *IBM*, *Cisco*, *Dell* and *Red Hat* – build private solutions.

Public Cloud Models
The cloud infrastructure can also be used by the general public (Mell and Grance 2011). Companies, academic or government organizations, or a combination of these, own, manage, and operate the cloud infrastructure. A public cloud generally provides a selection of services simultaneously for all users (in the form of, e.g., business processes, business practices, applications, and infrastructures). Such services are provided on the basis of usage-dependent payment via the Internet. One of the most significant distinguishing features of a public cloud is that a cloud customer can neither technically, nor contractually, influence which other parties use the cloud service. Thus, cloud customers (unknowingly, in terms of extent and scope) share the underlying infrastructure, which is, however, completely abstracted from the application layer. Examples of public clouds include *ESDS's eNlight Cloud*, *Amazon Elastic Compute Cloud (EC2)*, *IBM's Blue Cloud*, *Sun Cloud*, *Google App Engine* and *Windows Azure Services Platform*.

Community Cloud Models
The cloud infrastructure is used exclusively by a group of people or organizations who have similar demands of the cloud service, such as sharing the same mission, security requirements, policy, or compliance considerations (Mell and Grance 2011). One or more of the community members, third parties, or a combination of these parties own, manage, and operate the cloud infrastructure.

For example, IBM offers its *Federal Community Cloud* to federal government organizations as part of the company's dedicated federal data centers service, which provides secure certified computing capabilities. Cloud customers can decide on community solutions that *Google*, *Red Hat*, *IBM*, *Microsoft,* or others provide.

Hybrid Cloud Models
The hybrid cloud infrastructure consists of a combination of two, or more, of the above-mentioned models (in particular public and private cloud). The individual infrastructures remain unique entities, but are connected by standardized or proprietary technologies. This allows the transfer of data and applications between the connected infrastructures. The purpose of this mixed form of services is to create a solution that best meets the concrete requirements of each company. For example, a cloud customer could run a mission-critical workload within a private cloud, but use the database services of a public cloud provider for non-critical data.

Virtual Private Cloud Models

The term "virtual private cloud" was first used widely by *Amazon Web Services* when its new product, "*Amazon VPC*", was introduced. In the virtual private cloud model, the cloud provider supplies the underlying infrastructure exclusively to a single organization, which could include a number of users (e.g., business divisions) (Dillon et al. 2010). Access to the cloud service can be realized using a Virtual Private Network. The cloud infrastructure remains the cloud provider's property, and is thus operated and managed by the cloud provider. A virtual private cloud ensures that a cloud customer has complete control of the virtual resources.

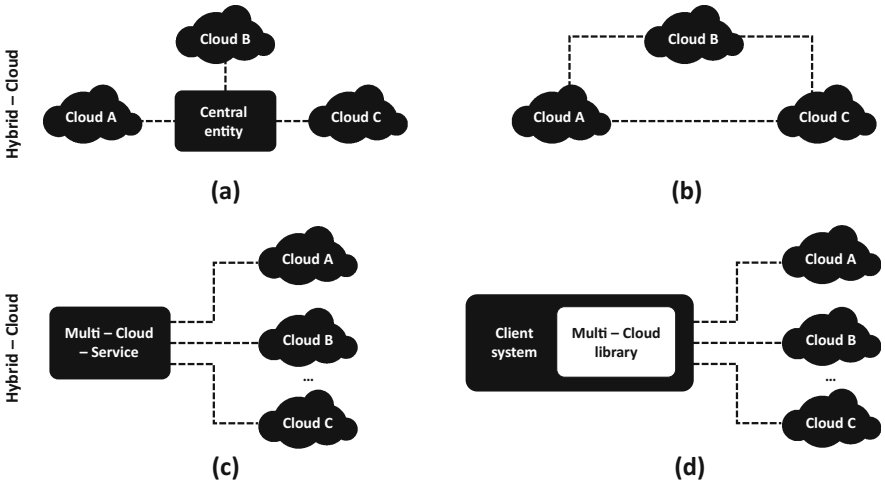

Fig. 7.2 Illustrative types of hybrid and multi-clouds (adapted from Grozev and Buyya (2014))

Multi-Cloud Models

If cloud services of different cloud providers are aggregated and combined, this is referred to as a multi-cloud (Grozev and Buyya 2014). Here, cloud providers can voluntarily connect their cloud infrastructures and services with other cloud providers, or a cloud broker enters the market, aggregating different cloud services from (different) cloud providers. The distinction between a multi-cloud and a hybrid cloud inconsistent and difficult to articulate. In contrast to a hybrid cloud, in which the cloud infrastructures are usually connected and work together (orchestrated), in multi-clouds, only certain cloud service components are specifically used by another cloud provider. For example, a multi-cloud provider could perform the computing and network operations in an *Amazon Web Service* cloud, while storage is performed solely by the *Microsoft Azure* cloud. Fig. 7.2 illustratively represents hybrid and multi-cloud scenarios.

7.1.6 Differences Between Related Concepts

IT Outsourcing vs. Cloud Computing
IT outsourcing refers to transferring some or all of the IT related decision-making rights, business processes, internal activities, and services to external providers (Böhm et al. 2011). Cloud computing is an evolution and a specific form of IT outsourcing, and thus shares common characteristics, providing similar benefits to customers (Schneider and Sunyaev 2016). For instance, customers perform IT outsourcing or adopt cloud services to reduce costs, to minimize risk by using third party vendors, and to outsource control and application management, as the third party carries out these processes. Customers share similar security and privacy risks in IT outsourcing and cloud computing contexts, as either way, the data is handled by a (potentially untrustworthy) third party. From a consumer's perspective, the main distinguishing characteristic of cloud computing compared to traditional IT outsourcing, is the flexible deployment of virtual and asset-free resources and services (Böhm et al. 2011). This allows the implementation of flexible, pay-per-use business models. Cloud providers are able to provide existing customers with a new kind of flexibility, and they can reach entirely new customer groups by offering new services and business models. In addition, the cloud computing model allows for the modification and extension of existing services without large upfront investments. The cloud provider *JungleDisk*, for example, used the IaaS services of *Amazon* to build a new business model that offers end-users easy to use storage services.

Further peculiarities of cloud computing distinguish it from IT outsourcing (Schneider and Sunyaev 2016). For instance, compared to traditional IT outsourcing, cloud computing induces a shift in task responsibilities during decision processes and self-service procurement; provides standardized services with a narrower scope; enables new scenarios of outsourcing and governance arrangements; and, uses short-term usage-based contracts.

IT Outsourcing
IT outsourcing refers to transferring some or all of the IT related decision-making rights, business processes, internal activities, and services to external providers (Böhm et al. 2011).

Application Service Provision vs. Cloud Computing
Application service provision is a form of outsourcing in which organizations rent packaged software and associated services from a third party (Bennett and Timbrell 2000). In cloud computing, SaaS is the closest corresponding service model to application service provision. Researchers argue that SaaS emerged as an advanced form of application service provision (Benlian and Hess 2011), that cloud computing has the same key attributes as the more traditional application service provision model, and that it exposes users to the same risks (Schwarz et al. 2009). Further, application service provision and SaaS share similar business and pricing models (Weinhardt et al. 2009). However, early providers offering application services

failed due to insufficient bandwidth and computing power (Susarla et al. 2003). Finally, cloud computing offers also more diverse service models than providing applications as a service.

Application Service Provision
Application service provision is a form of outsourcing in which firms rent packaged software and associated services from a third party (Bennett and Timbrell 2000).

Grid Computing vs. Cloud Computing
Grid Computing enables resource sharing and coordinated problem solving in dynamic, multi-institutional, and large-scale geographically-distributed virtual organizations (Foster et al. 2001). Grids provide a distributed, peer-to-peer computing paradigm or infrastructure that spans multiple virtual organizations; each virtual organization consists of either physically distributed institutions or logically related projects/groups (Foster et al. 2008). The goal of such a paradigm is to enable federated resource sharing in dynamic, distributed environments. One famous example of grid computing is the SETI project. The SETI application looks for radio signals or other forms of communications in space. Given a high demand for computing resources, SETI developed a grid computing middleware where the application can be executed over multiple computers that are connected to the Internet. Individuals can then download and install the middleware, which used idle computing powers at the disposal for SETI.

Clouds and grids share a great deal in their vision, architecture, and technology (Foster et al. 2008). The evolution of cloud computing resulted from a shift in focus from an infrastructure that delivers storage and computing resources (as is the case in grids) to one that is economy-based and aims to deliver more abstract resources and services (as is the case in clouds). The vision of grids and clouds remains the same, namely to reduce the cost of computing, increase reliability, and increase flexibility by transforming desktop computers and internal IT infrastructures to a facility operated by a third party. Clouds and grids share functions, but also some common challenges, such as a need for managing large facilities, for defining methods by which customers discover, request, and use resources provided by the central facilities, and a need for implementing the often highly parallel computations that are executed on those resources. Nevertheless, clouds and grids also differ in various aspects such as security, programming model, business model, computing model, data model, applications, and abstractions (Foster et al. 2008). In a cloud-based business model, a customer usually pays the cloud provider on a consumption basis, very much like paying for electricity, gas, and water. The model relies on economies of scale in order to drive prices down for customers and profits up for providers. In contrast, the business model for grids is project-oriented, which involves that users or a community have a certain number of service units (i.e., CPU hours) they can spend to achieve the project objectives. Further, clouds mostly comprise dedicated data centers that belong to the same organization. Within each data center hardware

and software configurations, as well as supporting platforms, are generally more homogeneous. Grids, however, build on the assumption that resources are heterogeneous and dynamic, so that each grid site might have its own administration domain and operation autonomy.

Grid Computing
Grid Computing enables resource sharing and coordinated problem solving in dynamic, multi-institutional, and large-scale geographically-distributed virtual organizations (Foster et al. 2001).

7.2 Essentials to the Provision of Cloud Services

7.2.1 Essential Cloud Technologies

Cloud services use a set of technology components that enable key features and characteristics of cloud computing (Singh and Chatterjee 2017). This section defines such technologies that help with understanding how these technologies are used to build cloud infrastructures.

Broadband Network and Internet Technology
The Internet allows customers to access remote cloud resources in order to support ubiquitous network access (Singh and Chatterjee 2017). Cloud computing's inherent network access requirement creates a built-in dependency on the Internet or private networks. The Internet is established and deployed by the Internet Service Provider (ISP). Each ISP can freely choose, manage, and add another ISP to their networks.

Data Center Technology
Data center technology contains multiple technologies and components that are typically connected to one another (Singh and Chatterjee 2017). The data center is built with standardized commodity hardware and modular architecture, multiple aggregation, unique building blocks that provide scalable, incremental growth of the services and reduce cost of investment in the cloud. The data center has both physical and virtualized IT resources. Physical IT resources include networking systems, servers, and equipment. Virtualized IT resources are placed in the virtualization layer that is operated and managed by the virtualization platform. Data center automation is also becoming increasing important because automating the bulk of the data center operations, management, monitoring and maintenance tasks that otherwise are performed manually by human operators increases efficiency and reduces costs.

Virtualization Technology
Virtualization refers to the conversion process that translates physical IT resources into virtual IT resources, such as virtual machines (Singh and Chatterjee 2017). The virtual IT resources include servers, storage, network, and computing power. Virtualization conceals the physical characteristics of IT resources and instead, presents an abstract, emulated virtual machine (Marston et al. 2011). A hypervisor

manages virtualized IT resources by running them on the physical host and providing coordination capacities (Bayramusta and Nasir 2016). The virtualized computing infrastructure is much better utilized than a purely physical infrastructure, which leads to lower upfront and operational costs. Virtual machines behave like an independent system, but unlike a physical system they can be started, closed down, and configured on demand; they can also be maintained and replicated very easily. Modern virtualization techniques allow cloud services to transfer virtual machines from a server with less space to another that has more space. This ability to re-allocate storage and computing power increases the flexibility and scalability of computing operations.

Container Technology

Containers refer to a similar but lighter virtualization concept (see Fig. 7.3). Containers are less resource and time-consuming, thus they have been suggested as a solution for improved interoperable application packaging in the cloud (Pahl 2015). They offer a logical packaging mechanism in which applications can be abstracted from the environment in which they actually run. Such detachment allows for easy and consistent deployment of container-based applications, regardless of whether the target environment is a private data center or a public cloud. Containerization provides a clean separation of concerns, as developers focus on their application logic, while IT operations teams focus on deployment and management of containers without worrying about the application details, such as versioning and specific configurations of the application. Although virtual machines and containers are both virtualization techniques, they solve different kinds of problems. Containers are tools for portably delivering software that aims at greater interoperability while still utilizing software virtualization principles. Virtual machines, on the other hand, take care of hardware allocation and management, with machines that can be turned on/off. Containers can replace virtual machines if the allocation of hardware resources is done using containers that divide workloads between clouds.

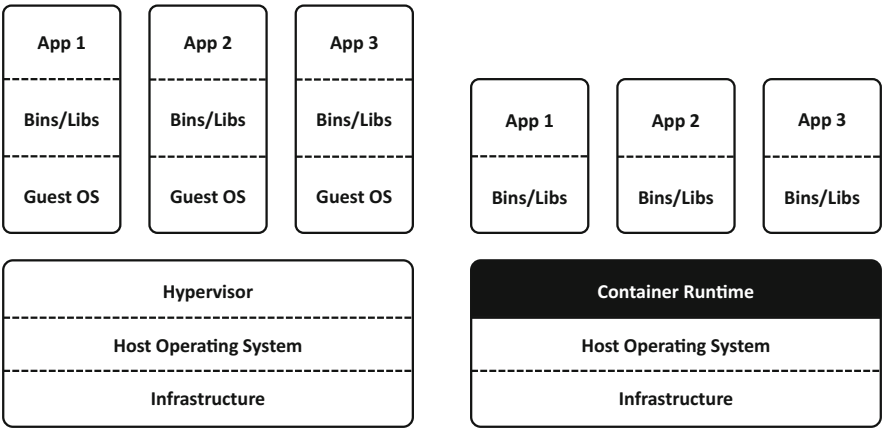

Fig. 7.3 Comparison of virtual machines and containers

Multi-Tenant Technology
Multi-tenant technology enables multiple users to access the same application logic simultaneously (Singh and Chatterjee 2017). Multi-tenant applications ensure that each tenant has a separate own view of the application, and tenants are not allowed to access data and applications provided to other tenants. Tenants can individually manage their features of the application. Such features are the user interface, business process, data model, and access control. The most important challenges multi-tenant applications face relate to usage isolation, data recovery, data tier isolation, data security, application upgrades, system scalability, and metered usage.

Service Technology
Service technology is the basic foundation of cloud computing, used for creating "as-a-service" cloud delivery models (Singh and Chatterjee 2017). Web service, REST service, service agents, and service middleware are basic technologies for building the cloud-based environment.

7.2.2 Cloud Service Stack

Cloud computing is based on a network of interrelated services on different levels of abstraction that form a layered architecture, which is also referred to as the 'cloud stack'. This layered architecture mirrors the various cloud computing service models (i.e., IaaS, PaaS, SaaS layers). The cloud stack helps to delineate responsibilities between cloud provider, cloud customers, as well as sub-providers. Table 7.2 schematically outlines the potential responsibilities. Depending on the cloud services' architecture, deviations to this simplified representation could occur in practice. Moreover, a cloud service is not limited to a specific operational environment or layer of the cloud stack. The layers can be spread out across all levels and, in the sense of the networked service structure, can be operated simultaneously by legally independent sub-providers. Thus, different parties and a combination of parties (e.g., cloud providers and sub-providers) can be responsible for platform security.

The following paragraphs will exemplify the separation of responsibilities across the three main cloud service models SaaS, PaaS, and IaaS. When using a SaaS model, a cloud customer generally is unable to make any changes to the cloud service. A cloud customer is only able to adjust certain configurations or settings on related cloud applications, such as turning certain features on or off, or customizing graphical user interfaces. Notably, the cloud customer is responsible for using the cloud application securely and in compliance with existing regulations. The SaaS provider's main responsibilities include the development, operation, maintenance, and administration of the software application, as well as ensuring software security. The remaining cloud layers can be operated by the SaaS cloud provider (full-stack provider) or outsourced to a sub-provider.

Table 7.2 Illustration of the different layers of the cloud service stack

Party	IaaS	PaaS	SaaS	Description of the layer
Cloud customer	Secure application usage			The cloud customers are responsible for secure usage of the application.
	User specifics			User-specific settings or configurations of used applications.
	Application		Application	Offered software solutions.
	Software security		Software security	Mechanisms to increase the security of provided applications.
	Administration and support		Administration and support	Administration of the software as well as receipt and handling of support inquiries from the cloud customer.
	Operating system	Operating system	Operating system	Basic software for operation of the application.
	Runtime environment	Runtime environment	Runtime environment	The runtime environment carries out applications for which the runtime environment is suitable.
	Database	Database	Database	Software for the management and structuring of data.
	Platform security	Platform security	Platform security	Mechanisms to increase the security of provided platforms.
	Administration and support	Administration and support	Administration and support	Administration of the provided platform as well as receipt and handling of support inquiries from the cloud customer.
	Virtual machines	Virtual machines		Virtual representation of computer resources like, for example, servers or CPUs.
Cloud provider	Virtualization	Virtualization		Mechanisms for the creation and management of virtual machines.
	Computing	Computing		Components to process data in the cloud service.
	Storage	Storage		Mechanisms for the storage of data.
	Network	Network		Mechanisms for the transfer of data.
	Infrastructure security	Infrastructure security		Mechanisms to increase the security of provided resources.
	Administration and support	Administration and support		Administration of the provided infrastructure as well as receipt and handling of support inquiries from the cloud customer.
	Hardware			The physical hardware for operation of the cloud service.
	Premises, housing and equipment			The physical set-up of the cloud service.
	Network			Physical connectivity of the data center.
	Data center security			Mechanisms to increase the security of the data center, including parties responsible for the housing, with security staff and physical security systems.

Key:

Cloud customer
SaaS core business
PaaS core business
IaaS core business
Full stack provider or sub-provider

In the case of a PaaS, cloud customers deploy and run their own applications on a cloud platform. Thus, a cloud customer is responsible for the development and operation of the applications, and has the possibility of some control over the hosting environment settings. Additionally, a cloud customer is responsible for securing the application, for example, to prevent cross-site scripting or software flaws. The PaaS provider's main responsibilities are the development, operation, and administration of the platform, caring for components such as runtime software execution stack, databases, and other middleware components, as well as ensuring platform security. PaaS cloud providers typically also support the development, deployment, and management of the cloud customer by providing tools, such as integrated development environments (IDEs), software development kits, deployment, and management tools (Liu et al. 2011). The remaining cloud layers can be operated by the PaaS cloud provider (full-stack provider) or are outsourced to a sub-provider.

The cloud provider of the IaaS is responsible for securely virtualizing and providing the necessary physical resources. The IaaS cloud provider typically has control over the hardware and cloud software that makes the provision of these infrastructure services possible. For example, it oversees the physical servers, network equipment, storage devices, host operation systems, and hypervisors for virtualization (Liu et al. 2011). A cloud customer is responsible for provisioned virtual machines and the applications, databases, operating systems, and runtime environments. In addition, the cloud customer takes responsibility for software and platform security. The required physical hardware, housing, and equipment can be acquired and managed by an IaaS provider (full-stack provider) or can be obtained from a sub-provider.

Depending on the specific design of the cloud stack and the service level agreements made with cloud customers, different responsibilities could arise. Regarding further literature, readers could refer to the *NIST 'Cloud Security Reference Architecture'*, which provides a detailed exposition of sharing responsibilities in various service models (NIST Cloud Computing Security Working Group 2013; Liu et al. 2011).

7.3 Chances and Challenges of Cloud Computing

7.3.1 Reasons to Move into the Cloud: Benefits and Opportunities for Organizations

Essentially, cloud computing enables individuals and organizations to access IT resources on-demand, from any device, at any time, and as a measured service. Due to its inherent characteristics (e.g., on-demand self-service, resource pooling, elasticity, and extensibility), cloud computing enables persons and organizations to achieve diverse benefits and opportunities. More importantly, cloud computing

exhibits transformative mechanisms that enable truly innovative services and business models, that ultimately lead to fundamental and large-scale innovations that benefit individuals, organizations, markets, and societies (Benlian et al. 2018). In the following sections, major benefits for organizations moving to the cloud will be presented, as well as the transformative mechanisms of cloud computing.

Low Entry Barriers
One of the cited benefits of cloud computing is a lowered entry barrier to large data centers that provide unique services for niche markets previously inaccessible due to high capital costs (Arasaratnam 2011; Venters and Whitley 2012). This is facilitated by customers sharing a significant amount of infrastructure which reduces costs per customer. This allows for economies of scale from both a service management and computing resource perspective. Cloud computing, therefore, enables small and medium-sized organizations to enter new markets quickly, particularly in areas like business intelligence which previously required significant IT investment (Venters and Whitley 2012). For emerging countries, cloud computing offers yet another opportunity to 'leapfrog' to advanced technology by connecting to cloud providers outside their countries (Venters and Whitley 2012). Some cloud computing providers are using the advantages of a cloud platform to enable IT services in countries that would traditionally have lacked the resources for widespread deployment of IT services (Marston et al. 2011).

Pay-as-you-go
Cloud services have brought about the reclassification of IT from an expensive 'capital expenditure' (requiring large upfront investment which could be difficult to raise) to a pay-as-you-go 'operating expenditure' (Venters and Whitley 2012). For example, cloud services enable universities and education institutions to provide computing laboratories to students and bulk processing services for research on a pay-per-use basis (Sarkar and Young 2011; Venters and Whitley 2012). Customers can thereby optimize the use of IT resources by transferring fixed costs to variable cost (Baldwin et al. 2001). This approach affects the customers' cash flow, with small, stable cash flows occurring rather than large periodic payments for licenses, maintenance contracts, and upgrades (Benlian and Hess 2011).

Access to Leading Edge IT Resources, Skills, and Capabilities
Access to leading edge IT resources has been shown to be one of the main indicators of IT outsourcing success, and an important driver of outsourcing decisions (Gonzalez et al. 2009; Benlian and Hess 2011). Cloud providers benefit from economies of scale by using a multi-tenant platform architecture, and consolidating and virtualizing its data centers. Further, the provider also benefits from learning curve effects when the delivery of cloud services via the Internet is professionalized (Benlian and Hess 2011). As a result, cloud customers benefit from economies of skill by leveraging the skills, resources, and capabilities that the service provider offers. These specialized capabilities (e.g., access to the latest technologies and IT-related know-how) could not be generated internally if the IT resources were delivered in-house via an on-premises model (Kern et al. 2002). Other business

benefits of cloud can include a reduced demand for skilled labor where skills shortages exist (Luftman and Zadeh 2011).

Quality Improvements
Customers frequently rely on cloud services to improve the quality and productivity of IT services by outsourcing to a third-party vendor (Benlian and Hess 2011; Schneider and Sunyaev 2016). They expect providers to incorporate industry's best practices and total quality management procedures, such as lean management concepts (Benlian and Hess 2011), and to aim for various quality improvements, such as a faster response time to end-users, or higher-quality user interfaces and features (Bajaj 2000). In general, cloud services are assumed to provide better reliability and availability due to a robust architecture. For example, when Netflix migrated to the cloud, not only the service quality was improved, but Netflix product itself has continued to evolve rapidly, incorporating many new resource-hungry features and relying on ever-growing volumes of data. Cloud-based business relationships between providers and customers, which are commonly based on key performance indicators (such as end-user response time or net uptime), allow for increased (outcome) measurability of service quality. Such relationships also enable clear contractual specifications regarding adequate service levels (e.g., with explicit provisions, which include fines, penalties, and even contract termination) (Benlian and Hess 2011). These characteristics in turn allow for higher transparency than that of an internal IT department, and could translate into stronger cloud provider discipline and better service quality.

Cost Savings
Cloud customers seek to achieve cost advantages by relying on cloud services, because external cloud providers can provide IT functions, such as managed application services, at lower costs than customers can (Benlian and Hess 2011; Venters and Whitley 2012). This ability is due to the providers' specialization and achievement of economies of scale and scope. Since the cloud business model runs a shared infrastructure and provides multiple users with a single instance of a highly standardized software service, it can be based on an extremely scalable and cost-efficient platform. Such improved economics in the provision of cloud-based resources can be passed on to customers, who can benefit from the lower total ownership costs. Treating IT as an operational expense helps to reduce the upfront costs in corporate computing quite dramatically. For example, many of the promising new Internet startups like *37 Signals*, *Jungle Disk*, *Gigavox*, and *SmugMug,* were realized with investments in information technology that are orders of magnitude lesser than those that would have been required just a few years ago.

Focus on Core Capabilities
In general, outsourcing enables organizations to focus on their core business, because they can free up resources, which can be used more productively in areas that create value (Benlian and Hess 2011; Schneider and Sunyaev 2016). This refocusing is possible by completely shifting the responsibility for developing,

testing, and maintaining the outsourced resources and the underlying infrastructure to the cloud provider. This shift will not only relieve line managers of the task of coordinating a large IS department, but will also eliminate IT staff members' routine support activities (Gonzalez et al. 2009). The staff can then dedicate their time to more strategic activities to determine how the organizations' IT can add value to the business.

Greater Flexibility and Elasticity
Using cloud services provides a great degree of flexibility regarding the utilization of easily scalable and on-demand provision of IT resources (Benlian and Hess 2011). Customers can quickly scale resources both up and down and pay as you go based on the amount customers actually use. This flexibility makes it easier for organizations to respond to business-level volatility, because the cloud provider handles fluctuations in IT workloads. In this regard, a customer can leverage a cloud provider's capacity to adapt to change. Since the computing resources are managed through software, they can be deployed very quickly as new requirements arise (Marston et al. 2011).

Reduced Time to Market
When services are sourced externally, cloud computing enables customers to deliver their products or services to the market faster (Seethamraju 2013). Cloud providers afford almost immediate access to hardware resources, with no upfront capital investments for users, which leads to a faster time to market in many businesses (Marston et al. 2011).

Lower IT Barriers to Innovation
Cloud computing can lower IT barriers to innovation, as the many promising startups testify, like *TripIt* (for managing travel arrangements) or *Mint* (for managing personal finances) (Marston et al. 2011). Cloud computing also makes possible new classes of applications, and delivers services that were not possible before. Examples include: (1) interactive mobile applications that are location, environment, and context-aware and that respond in real time to information provided by human users, nonhuman sensors (e.g., humidity and stress sensors within a shipping container), or even by independent information services (e.g., worldwide weather data); (2) Parallel batch processing, that allows users to take advantage of huge volumes of processing power to analyze terabytes of data for relatively small periods of time, while programming abstractions like *Google's MapReduce* or its open-source counterpart *Hadoop* make the complex process of parallel execution of an application over hundreds of servers transparent to programmers; (3) Business analytics that can use the vast body of computer resources to understand customers, buying habits, supply chains, and so on, from voluminous amounts of data; (4) Extensions of compute-intensive desktop applications that can offload the data crunching to the cloud, leaving only the rendering of the processed data at the front-end, with the availability of network bandwidth reducing the latency involved.

7.3.2 *Cloud Computing's Transformative Mechanisms*

Whereas cloud computing has traditionally been viewed as a means to generate IT value, e.g., economic advantages through the deployment and use of IT (Schneider and Sunyaev 2016), cloud computing exhibits transformative mechanisms that enable truly innovative services and business models, ultimately leading to fundamental and large-scale innovations that benefit individuals, organizations, markets, and societies (Benlian et al. 2018). In the following sections, we describe three transformative mechanisms of cloud computing, namely decoupling, platformization, and recombination.

Decoupling
Decoupling describes a process in which one element of a system becomes an independent service with a defined service interface so that internal changes in this service will not disrupt the functioning of other dependent elements (Tiwana et al. 2010). In cloud computing, decoupling started on the infrastructure level (Benlian et al. 2018). Enabled by virtualization techniques, application systems have become independent of their underlying physical resources. For example, the detachment of large resource-intensive applications run on virtualized layers that utilize multiple physical instances. Virtualization allows for elastically scaling these resources up and down—a key characteristic of cloud computing. Cloud computing also has exploited and spurred on the increasing decoupling on the component level. While modularization and decoupling has long been a trend in software engineering in both monolithic and distributed systems, widespread web service protocols (e.g., SOAP, REST) have increased the level of decoupling and made the reuse of third-party services a more common practice. Applications use functionality and resources from remote services, e.g., for retrieving geolocation information from *Google Maps*. On the cloud application level, decoupling has led to a greater separation between those who use applications and those who provide them—a logical continuation of the outsourcing trend (Schneider et al. 2018; Schneider and Sunyaev 2016). Traditional IT outsourcing arrangements were single-tenant with IT resources dedicated to a specific user organization (Yoo et al. 2010). Cloud computing services are provided for use by multiple tenants with a postulate of minimal service provider interaction, which further decouples the provider-user relationship.

Decoupling
Decoupling describes a process in which one element of a system becomes an independent service with a defined service interface so that internal changes in this service will not disrupt the functioning of other dependent elements (Tiwana et al. 2010).

Platformization

The second mechanism, platformization, builds on decoupling and characterizes the process in which an entity (a provider organization) creates access and interaction opportunities centered around a core bundle of services (the platform) within an ecosystem of customers, complementors (i.e., cloud service partners), and other stakeholders (Tiwana et al. 2010; Cusumano 2010). On all cloud computing layers, new players have emerged that develop capabilities by providing infrastructure resources, applications, and component services, leading to the commercial reality of IaaS, SaaS, and PaaS (Benlian et al. 2018). Customers harness cloud platforms to gain access to easily connectable services, and are able to satisfy their unique needs due to cloud platforms providing greater variety, service quality and business flexibility than otherwise. On the resource layer, *Amazon*, for example, made unused virtualized computing resources available to the external market with offerings like the *EC*2 (the elastic cloud) and swiftly became a market leader for IaaS. From a customer perspective, the large data centers of public cloud providers enabled practically infinite possibilities for scaling computing resources. On the component level, PaaS providers, such as *Google App Engine*, *IBM Bluemix*, and *Heroku*, bundled online development and execution environments with an increasing number of services, offering developers who design cloud-based applications from reusable components rather than building software from scratch. On the application level, cloud-native software vendors like *Salesforce.com* specialized in specific enterprise software segments, and thereby earned tremendous success since they provided cost-effective solutions to markets that would otherwise not have been able to afford these systems. These IaaS, PaaS, and SaaS providers have quickly turned into platform businesses and built up an ecosystem of customers, complementors (i.e., cloud service partners), and third parties that exchange knowledge and other resources with or via the platform (Grover and Kohli 2012). Cloud platform players have gained the critical size required for their ecosystems to flourish and to co-create new services, which ultimately benefit cloud customers.

Platformization
Platformization characterizes the process by which an entity (a provider organization) creates access and interaction opportunities centered around a core bundle of services (the platform) within an ecosystem of customers, complementors (i.e., cloud service partners), and other stakeholders (Tiwana et al. 2010; Cusumano 2010).

Recombination

The third transformative mechanism, recombination, refers to the process in which innovation potential is generated by combining cloud services and integrating them with other transformative technologies within and across platform ecosystems (Benlian et al. 2018). Recombination takes place on all three service models. For

example, IaaS platforms allow third-parties (complementors) to offer basic resources (e.g., pre-configured virtual instances, Internet of Things utilities) on marketplaces where users can choose from myriads of configuration options that will fulfill their specific needs. On the component level, PaaS has opened the door to the application programming interface (API) economy (Tan et al. 2016). Not only do PaaS platforms like *Amazon AWS*, *Microsoft Azure*, and *IBM Bluemix* allow complementors to offer value adding components (e.g., analytics, artificial intelligence, and blockchains); developers also integrate cross-platform via APIs with other open services from the programmable web to provide best-of-breed solutions that were formerly unthinkable. On the application level, all major software vendors aim to cultivate marketplaces with apps from third-party developers that offer similar innovation opportunities that the platform alone could never seize. *SAP App Center*, and *Microsoft Dynamics Marketplace* are only two examples from the large body of enterprise software vendors that have successfully created the app marketplace in the enterprise software world.

Recombination
Recombination refers to the process in which innovation potential is generated by combining cloud services and integrating them with other transformative technologies within and across platform ecosystems (Benlian et al. 2018).

7.3.3 The Downside of Cloud Computing: New Risks and Challenges

Although there are many benefits to cloud computing, it is not without its own risks and challenges. Cloud computing differs from previously studied products and services in the sense that it introduces ongoing uncertainty in the relationship between the provider and the customer (Trenz et al. 2018). Although customers depend on the cloud provider at all times, they typically lack personal interaction with the provider and have only limited information about the providers' qualities, intentions, and (future) actions. The sections below present several major risks and challenges concerning the use of cloud services. Specifically, the following subchapter discusses security and data protection issues in more detail.

Multitenancy
Although multitenancy affords cloud customers unprecedentedly low prices, security, privacy and compliance considerations need to be taken into account (Arasaratnam 2011). Depending on the layer of cloud service being provided, appropriate security controls should be employed. This can include host intrusion detection systems, hypervisor based security agents, host firewalls, and many others. Some cloud providers will enforce particular controls on their customers to ensure a minimal level of security. Others will allow their clients full flexibility. Cloud

customers who have concerns regarding the security of their workload in a multitenant environment should consult their provider regarding their security standards. If the provider's standards are deemed insufficient, the customer should address this with compensating controls, or defer to an alternative provider.

Loss of Control
Traditional computing models have permitted a high degree of control over compute resources (Arasaratnam 2011). Cloud computing, by virtue of abstraction, prevents the customer from having the same level of influence over the computing resource. In this regard, the cloud customer loses the administrative power, as well as operational and security control over the cloud system. In addition, cloud providers can outsource or sub-contract services to third-parties (i.e., sub-providers) that might not offer the same commitment as the provider does. Customers can only assess the provider's business operations and capability to fulfill transactions through visiting their website and consuming the provided online service itself. Hidden actions of the provider might never be detected because the customer cannot continuously monitor providers' actions. For instance, cloud customers can never fully assess whether cloud providers disclose customer's information to third parties without their consent (Trenz et al. 2018). In other cases, there can be a significant time lag before customers recognize providers' concealed actions, e.g., through media disclosure.

Location Intransparency
Location intransparency is a concern linked to social or regulatory contexts rather than technological issues (Ahmed and Litchfield 2018). If data from one region (e.g., nation and jurisdiction) is transferred to any other region, there is no guarantee that the data is being treated in the same way as the source would do. This includes the level of security, as well as retention and processing of data. Ideally, regulations that address data management should be consistent across jurisdictions. However, different countries and regions have different requirements regarding how its citizens' information should be handled (Arasaratnam 2011). In some areas, such as the European Union (EU), there are specific requirements regarding EU residents' data protection. In other areas, such as the United States, there are directives regarding protected health information, such as the Health Insurance Portability and Accountability Act. The challenge is that if the customer doesn't know where all the cloud provider's assets reside, it is difficult to determine with which legislation the customer has to comply. Further, if the cloud provider has multiple data centers worldwide, in many instances it is impossible to tell where in the world a particular customer's data set might be at any one point in time.

Lack of Availability
Customers have certain expectations about the service level to be provided once their applications are moved to the cloud (Voorsluys et al. 2011). These expectations include availability of the service, its overall performance, and what measures are to be taken when something goes wrong in the system or with its components. Typically, these expectations are settled in the SLA with the cloud provider. Yet,

cloud services face several threats regarding their availability. A nefarious user within the cloud can adversely affect the performance of or availability to other customers by attempting to over-consume resources (Arasaratnam 2011). A similar effect can result from customers using the significant elasticity of the cloud to launch a denial of service attack against other customers or organizations. Although most cloud providers have defense mechanisms against these attacks, law enforcement can be just as disruptive. Consider a recent case where law enforcement officials confiscated racks of servers from a data center due to alleged illegal activity by one of the customers. Unfortunately, due to the use of virtualization many tenants' information was confiscated at the same time. Hardware availability is another issue in cloud computing (Singh and Chatterjee 2017). If hardware resources are unavailable, it can lead to cloud outages that hamper the entire online business fraternity, thus causing considerable distress. For example, in late 2007, the cloud provider *Rackspace* suffered a 36-hour outage due to a damaged transformer outside their data center (Arasaratnam 2011). This significant issue affected all of their customers.

Compromised Ease of Use
Customers might be uncertain about the effort required to use the cloud service (Venkatesh et al. 2003). If the cloud service's computer interface is simple, easy to use, and does not require specialized skills, it is more likely to be perceived as useful (Lee and Wan 2010). Conversely, if the interface is difficult to understand and use, and requires specialized skills, it is likely to be perceived as less useful, in which case customers will be reluctant to adopt it. The degree of effort can be taken as a cue signaling that the system has been well tested and has the ability to work effectively. Such effort could also indicate that the provider is actually investing in the customer relationship (Gefen et al. 2003). Cloud services might exhibit a high degree of complexity, e.g., due to a broad range of functionalities, complex interface designs, or innovative capabilities.

Vendor Lock-In
Customers are confronted with uncertainty about whether they can leave the cloud service without incurring social or economic losses (Bhattacherjee and Park 2014; Trenz et al. 2013). Such losses are defined as switching costs, referring to one-off costs that customers associate with the process of switching from one provider to another (Burnham et al. 2003). There are different types of procedural switching cost, which are particularly relevant due to the low investment necessary to use a cloud service. Procedural switching costs comprise learning costs that arise from the time and effort of acquiring new skills or know-how required to use a new service effectively. This type of switching cost includes setup costs that represent the costs associated with the time and effort necessary for initiating a relationship with a new provider or setting up a new service. Further, switching costs for cloud services can be of a social nature (Jones et al. 2002), as is, e.g., reflected in the lost benefit of sharing files with other customers who use the same service. Likewise, a major

concern customers have is about having their data locked-in by a certain provider (Voorsluys et al. 2011). Users might want to move data and applications away from a provider that does not meet their requirements. However, in their current form, cloud computing infrastructures and platforms do not employ standard methods of storing user data and applications, and thus data portability is limited.

Limited Service Continuity
Another legitimate concern is centered on cloud providers: customers are uncertain whether the service (or particular service functions) will continue to be offered over time (Marston et al. 2011; ENISA 2012). As in any modern IT market, negative consequences resulting from, e.g., competitive pressure, an inadequate business strategy, provider acquisition, or lack of financial support, could put some providers out of business or at least force them to restructure their service portfolio offering. Subsequently, customers fear that in the short or medium term some widely used service functions, or even the complete cloud service, could be terminated. The impact of such a threat on the cloud customer is very understandable, because it could lead to a loss or deterioration of their performance in quality service delivery, as well as losses in financial and non-financial investment (i.e., learning costs).

7.4 Security and Data Protection in Cloud Environments

7.4.1 Security and Privacy Challenges Due to Essential Cloud Service Characteristics

Cloud services are an attractive alternative to traditional information technology for organizations in diverse industries (e.g., healthcare) due to cloud computing's essential characteristics (Gao et al. 2018; Thiebes et al. 2017). However, these features challenge contemporary approaches to security and privacy risk assessment (Benlian et al. 2018; Hentschel et al. 2018). For example, a multi-tenant and virtualized approach seems promising from a cloud provider's perspective in terms of profit, but this approach increases the risk of co-location attacks due to inappropriate logical and virtual isolation. Consequently, an increasing number of research and industry reports has focused on identifying and addressing security and privacy risks, including Internet, network, and access security issues, as well as risks regarding non-compliance with regulatory requirements (cf. Fernandes et al. (2014) for an excellent review of security issues in cloud environments). In the following sections, major challenges regarding the security and privacy of cloud services will briefly be presented.

Communication Issues
Communication with the cloud is performed via the Internet, which gives rise to many security risks during data transmission (Sehgal et al. 2011). Threats include

man-in-the-middle attacks, hardware-based attacks, browser and network vulnerabilities, as well as exploiting injection and protocol vulnerabilities. For example, man-in-the-middle attacks are performed by an intruder, who has access to packets moving in a network (Kouatli 2014). The intruder can interfere in communication between the cloud provider and customer, and would then be able to monitor and manipulate the traffic between them.

Data Breaches
A data breach is an incident in which sensitive, protected or confidential information is released, viewed, stolen, or used by an individual who is not authorized to do so (Top Threats Working Group 2016). The potential for data breaches increases exponentially in cloud service contexts, because cloud data centers comprise data from multiple customers (Whitman and Mattord 2011). A data breach can be the primary objective of a targeted attack, e.g., when organized crime seeks financial, health, and personal information to carry out a range of fraudulent activities. Likewise, competitors and foreign nations could be interested in proprietary information, intellectual property, and trade secrets. Activists might want to expose information that can cause damage or embarrassment. Data breaches could simply be the result of human error, application vulnerabilities, or poor security practices. Unauthorized insiders obtaining data within the cloud are a major concern for organizations.

Data Loss
Although customers outsource their data to the cloud because the cloud infrastructure is much more powerful and reliable, and has more capacity than their own devices or servers, customers still fear that their data can get lost (Burda and Teuteberg 2014; Park et al. 2016). From time to time outages and data loss incidents do occur in noteworthy cloud storage services (e.g., data loss incidents of storage providers *Linkup*, *Carbonite Inc.*, *Amazon's EC2* and *Google's Cloud*). Cloud services face a broad range of both internal and external threats to data integrity that can lead to data loss. Such threats include hardware or system malfunctions, human error, software corruption, or back-up failure (Subashini and Kavitha 2011). Similarly, data stored in the cloud can be lost due to the cloud provider accidentally deleting it, or worse, a physical catastrophe such as a fire or earthquake causing the permanent loss of customer data. Consequently, cloud providers typically operate multiple data centers to back data up in different locations.

Data Scavenging
Cloud servers typically reuse the storage space once the data has been properly utilized and sent to trash folders (Singh and Chatterjee 2017). In multi-tenant cloud environments, providers have to ensure that data used by a previous user is not available to the next user. In what is referred to as data scavenging, attackers could be able to recover removed data that, not having been destroyed properly, might still exist on the device, (Khan and Al-Yasiri 2016). The process of cleaning up or removing certain data units from a resource is known as data sanitization (Singh

and Chatterjee 2017). In distributed systems, data sanitization is a critical task in selecting the data destined for the trash, and for properly disposing of discarded data.

Insufficient Identity, Credentials, and Access Management
Cloud customers are often concerned about who can access their data in the cloud. Data breaches and successful attacks can occur due to lacking scalable identity access management systems, failure to use multifactor authentication, weak passwords, and a lack of ongoing automated rotation of cryptographic keys, passwords, and certificates (Top Threats Working Group 2016). In cloud service environments, implemented identity systems must scale to handle lifecycle management for millions of users. Consequently, identity management systems have to support immediate de-provisioning of access to resources when personnel changes, such as job termination or role change, occur. Multifactor authentication systems – smartcard or phone authentication, e.g. – are required for users and operators of a cloud service.

Cloud customers have to be aware that not only outsiders can gain access to their data, but that the provider itself is often able to access such data. Furthermore, in these cases the provider usually tries to conceal employees' access rights from the customers. In what is referred to as a malicious insider threat to an organization, the offender is a current or former employee, contractor, or other business partner who has or had authorized access to an organization's network, system, or data, and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems (Top Threats Working Group 2016). This threat mainly materializes due to lacking transparency and the IT services and customers working in a single management domain (Singh and Chatterjee 2017). An employee can gain a higher level of access, and using this, can penetrate and compromise the confidentiality of data and services. This can also result in an insider attacker attaining access to confidential data and affecting the cloud services' operation.

Insecure Interfaces and APIs
Cloud computing providers display a set of software user interfaces or APIs that customers use to manage and interact with cloud services (Top Threats Working Group 2016; Singh and Chatterjee 2017). The security and availability of general cloud services is dependent on the security of these basic APIs. From authentication and access control to encryption and activity monitoring, these interfaces must be designed to protect against both accidental and malicious attempts to circumvent policy. Further, organizations and third parties can build on these interfaces to offer value-added services to their customers. This introduces the complexity of the new layered API, and it also increases risk, because organizations could be required to relinquish their credentials to third parties in order to enable their agency. APIs and user interfaces are generally the most exposed part of a system, perhaps due to being

the only asset with an IP address available outside the trusted organizational boundary. Such assets will be acutely targeted for heavy attack. Thus, adequate controls protecting assets against Internet fraud are the first line of defense and detection. A successful attack on the cloud interfaces can result in root level access to a machine without initiating a direct attack on the cloud infrastructure (Hussain et al. 2017).

System Vulnerabilities
System vulnerabilities are exploitable bugs in programs that attackers can use to infiltrate a computer system with the purpose of stealing data, taking control of the system, or disrupting service operations (Top Threats Working Group 2016). Vulnerabilities within the components of the operating system (i.e. the kernel, system libraries, and application tools) put the security of all services and data at significant risk. This type of threat is nothing new, as bugs have been a problem ever since the invention of computers and thus became exploitable remotely when networks were created. With the advent of multitenancy in cloud computing, systems from various organizations are placed in close proximity to each other, and given access to shared memory and resources, creating a new surface for attacks.

Shared Technology Vulnerabilities
Cloud providers deliver their services in a scalable manner by sharing infrastructure, platforms, or applications (Top Threats Working Group 2016). Underlying components (e.g., CPU caches, GPUs, etc.) that comprise the infrastructure supporting cloud services deployment may not have been designed to offer strong isolation properties for a multitenant architecture (IaaS), re-deployable platforms (PaaS), or multi-customer applications (SaaS). This can lead to shared technology vulnerabilities that can potentially be exploited in all delivery models. A single vulnerability or misconfiguration can lead to a compromise across an entire provider's cloud. Flaws in hypervisors can allow one virtual machine to control or access information on another (Whitman and Mattord 2011).

7.4.2 Continuous Service Certification as Innovative Means to Ensure Security and Data Protection

A common strategy in reducing customers' security and data protection uncertainty that will also signal trustworthiness and adequate risk prevention, is the adoption of certification. This is particularly important for small- and medium-sized cloud providers because they cannot rely on a widely established high reputation in the market (Sunyaev and Schneider 2013; Lins et al. 2016a; Malluhi and Khan 2013). Certification is defined as a third-party attestation of products, processes, systems, or persons that verifies conformity to specified criteria (ISO 2004).

Certification
Certification is defined as a third-party attestation of products, processes, systems, or persons that verifies conformity to specified criteria (ISO 2004).

However, different existing cloud service certifications represent only a retrospective view of the fulfillment of technical and organizational measures when the relevant certificates are issued (Lins et al. 2016a; Lins et al. 2018). Typically, certification authorities evaluate adherence to certification criteria during a comprehensive audit performed once. Throughout the validity period of one to three years, certification deviations or breaches might not be detected until long after their occurrence because certification authorities validate certification adherence only via spot checks during annual surveillance audits. Thereafter, conventional certifications cannot support dynamic changes in the structure, deployment, or configuration of the cloud infrastructure that supports the provision of cloud services, such as the dynamic migration of data and software across different computational nodes in cloud federations (Krotsiani et al. 2015). In addition, a cloud provider can deliberately discontinue adherence to security and privacy requirements to achieve benefits, such as reducing required incident response staff to save costs (Lins et al. 2016a; Lins et al. 2018). Considering a highly dynamic cloud infrastructure, long validity periods of conventional certification could cause cloud customers to question the reliability and trustworthiness of issued certificates. This ultimately threatens the ability of cloud certification to prove adequate risk prevention (Lins et al. 2019b).

To address the juxtaposition of static certifications in an ever-changing and dynamic cloud service environment, researchers have started to investigate how the process of certifying cloud services can be innovated (Windhorst and Sunyaev 2013). These research efforts have resulted in innovative specifications of architectures and processes, as well as in continuous cloud service certification prototypes that allow certification authorities to continuously certify cloud services. Continuous cloud service certification (CSC) includes automated monitoring and auditing techniques, as well as mechanisms for transparent provision of certification-relevant information to continuously confirm adherence to certification criteria (Lins et al. 2016a). The process of continuous cloud service certification includes four major dimensions: (1) (semi-)automated data collection and transmission, (2) (semi-) automated data analysis, (3) up-to-date certification presentation, and (4) adjustment of processes (see Fig. 7.4).

Continuous Cloud Service Certification
Continuous cloud service certification includes automated monitoring and auditing techniques, as well as mechanisms for transparent provision of certification-relevant information to continuously confirm adherence to certification criteria (Lins et al. 2016a).

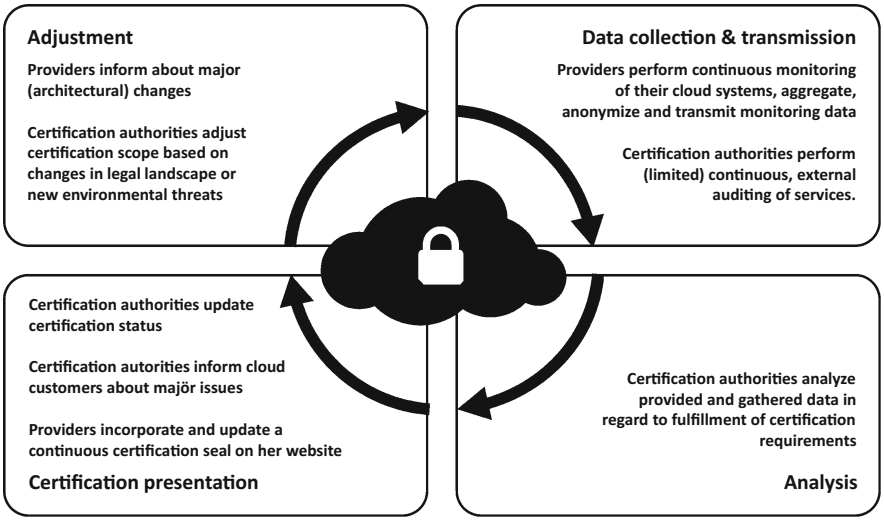

Fig. 7.4 Continuous certification process (adapted from Lins et al. (2016a))

In contrast to annual surveillance audits of conventional certification, automated data collection, and analysis of certification-relevant data, enable certification authorities to actively detect and investigate critical defects as they occur. Such procedures also enable immediate reaction to changes or events concerning a cloud service, and help certification authorities to adjust their certification reports based on an ongoing assessment of these defects, changes, and events (Lins et al. 2019b). Through timely detection and continuous assurance of certification adherence as required in highly dynamic cloud service environments, CSC can improve the trustworthiness of issued certifications. Consequently, in contrast to conventional certification, CSC considers the actual *status quo* of the cloud infrastructure when they assess certification adherence, and the CSC ultimately informs cloud customers on both infrastructure improvements (i.e., better service quality) or failures (i.e., data losses) through a transparent and up-to-date certification representation.

Two distinct but complementary types of CSC exist, namely test-based and monitoring-based certification (Anisetti et al. 2017; Kunz and Stephanow 2017; Lins et al. 2018; Stephanow et al. 2016). Test-based CSC methodologies are operated by certification authorities to collect certification-relevant information by directly accessing the cloud service infrastructure and testing cloud service components (Stephanow and Banse 2017). Typically, test-based certification techniques produce evidence by controlling some input to the cloud service and evaluating the output, such as calling a cloud service's RESTful API and comparing responses with expected results (Kunz and Stephanow 2017). Test-based CSC can be applied to verify the integrity of multiple cloud users' data (Wang et al. 2014), to assess data location (Doelitzscher et al. 2012), or to validate adherence to security criteria, among others (Stephanow and Khajehmoogahi 2017). In contrast, monitoring-

based certification techniques use data collected during monitoring as evidence collected from components involved in service delivery the cloud service is being provided (Kunz and Stephanow 2017). Implementing a CSC monitoring system enables cloud providers to (continuously) monitor their cloud infrastructures, and in the process collect and then transmit certification-relevant information to certification authorities (Lins et al. 2018; Lins et al. 2019b). For example, a prototypical monitoring-based CSC infrastructure (called *'CUMULUS'*) was developed to, for instance, verify database user identification that could validate certification criteria (Krotsiani et al. 2015).

Performing monitoring- or test-based CSC is beneficial for cloud providers, certification authorities, and cloud service customers alike (Lins et al. 2016b; Teigeler et al. 2018). Cloud providers receive ongoing third party expert assessment of their systems. This enables providers detect potential flaws and (security) incidents earlier than before, and can save costs due to successive service improvements. Certification authorities actively detect and investigate critical certification deviations as they occur, thus increasing certification reliability. These authorities can increase their auditing efficiency, achieve savings in their budgets, reduce operation times, and reduce operation fees by relying on automated auditing processes to reduce auditing time and errors. Similarly, compared to their manual predecessors, CSC is more cost-effective due to certification authorities being able to test larger samples and examine data faster and more efficiently. Cloud service customers can benefit from CSC as well. Typically, cloud environments lack control because cloud customers cede governance to cloud providers (ENISA 2012). CSC can counteract this lack of control by increasing transparency regarding providers' operations, and providing assurance regarding requirements (e.g., ensuring encryption, data integrity, and location), and so ultimately enhancing trustworthiness of cloud services. Informing customers in cases of critical certification violations or major security breaches is becoming increasingly important nowadays because organizations, individuals, and even societies and economies are highly dependent on cloud services during their day-to-day activities (Benlian et al. 2018).

Summary

As has been mentioned, cloud computing is an evolution of information technology and a dominant business model for delivering IT resources. With cloud computing, individuals and organizations can gain on-demand network access to a shared pool of managed and scalable IT resources, such as servers, storage, and applications. While cloud computing emerged from related computing paradigms, such as application service provision and grid computing, and it shares similarities with IT outsourcing and fog computing, there are several peculiarities that distinguish it from related paradigms. In particular, how individuals and organizations access and use IT resources has changed as a result of cloud computing's inherent characteristics, such as on-demand self-service, ubiquitous access, multitenancy, rapid

elasticity, pay-per-use, as well as various service models (i.e., SaaS, PaaS, IaaS) and deployment models (i.e., public, private, hybrid, community). Today, we rely heavily on cloud services in our daily lives, e.g., for storing data, writing documents, managing businesses, and playing games online. Additionally, cloud computing provides the infrastructure that has powered key digital trends such as mobile computing, the Internet of Things, big data, and artificial intelligence, thereby accelerating industry dynamics, disrupting existing business models, and fueling the digital transformation.

By harnessing a diverse set of technology components, including the Internet, data centers, virtualization, and multi-tenant technologies, cloud providers offer innovative cloud services that enable organizations to achieve a vast number of benefits and opportunities. These benefits comprise cost savings that accrue from, among other things, a pay-per-use procedure, greater flexibility, access to leading edge IT resources, skills, and capabilities, as well as from lower entry barriers to markets or innovative IT services, and from focusing on core capabilities. More importantly, cloud computing exhibits transformative mechanisms that enable truly innovative services and business models, ultimately leading to fundamental and large-scale innovations that benefit not only individuals and organizations, but also markets and societies. In particular, the transformative mechanisms enable a system to become an independent service with a defined service interface, which then can be bundled and recombined with other services to create a platform within an ecosystem of diverse stakeholders.

In spite of the above-mentioned benefits, customers also face a high degree of uncertainty when they use cloud services. In particular, customers lose control of outsourced IT resources, and face uncertainty about the location where their data is stored and processed. Likewise, customers can be concerned about whether the cloud service is available whenever it is needed, and will be easy to use. Cloud computing also challenges contemporary security and privacy risk assessment approaches. Security and privacy breaches, data loss, as well as insecure interfaces and APIs are just a few examples of security and privacy challenges that cloud providers face. A common strategy to reduce customers' security and data protection uncertainty and to signal trustworthiness and adequate risk prevention, is the adoption of certification processes. To address the juxtaposition of static certifications in an ever-changing and dynamic cloud service environment, researchers have recently started to investigate how to achieve continuous service certification that will increase certification reliability and trustworthiness.

Defying initial concerns, cloud computing has already become a critical IT infrastructure for almost every aspect of our everyday lives, and it will continue to transform the world we live in on multiple levels and in various ways. For example, cloud computing is essential for the growth of artificial intelligence, because most types of hardware do not have the capabilities to run artificial intelligence applications (i.e., machine learning or deep neural networks) efficiently. Also, while cloud computing improves processing speed and accuracy of artificial intelligence applications, artificial intelligence can also be used to operate and manage cloud computing in a more efficient way, as in workload scheduling in clouds. The increased

use of cloud computing further results in a greater demand for cloud professionals in the future. Students should become familiar with the cloud computing paradigm to compete in future job markets.

Questions

1. What are the main characteristics of cloud services?
2. How does PaaS differ from IaaS?
3. What are the major reasons for organizations moving into the cloud?
4. Which risks that cloud service customers face, have been addressed by cloud providers?
5. What are transformative mechanisms of cloud computing and how do they interact?
6. How can continuous cloud service certification overcome the issues associated with current certification processes?

References

Ahmed M, Litchfield AT (2018) Taxonomy for identification of security issues in cloud computing environments. J Comput Inf Syst 58(1):79–88

Anisetti M, Ardagna C, Damiani E, Gaudenzi F (2017) A semi-automatic and trustworthy scheme for continuous cloud service certification. IEEE Trans Serv Comput 10(1):1–14

Annette JR, Banu WA, Chandran PS (2015) Rendering-as-a-service: taxonomy and comparison. Procedia Comput Sci 50:276–281

Apps Run The World (2017) Cloud applications revenue from leading vendors worldwide in 2015 and 2016 (in million U.S. dollars). https://www.statista.com/statistics/475844/cloud-applications-revenues-worldwide-by-vendor/. Accessed 11 Dec 2018

Arasaratnam O (2011) Introduction to cloud computing. In: Halpert B (ed) Auditing cloud computing, a security and privacy guide. Wiley, Hoboken, NJ, pp 1–13

Bajaj A (2000) A study of senior information systems managers decision models in adopting new computing architectures. J AIS 1(1es):4

Baldwin LP, Irani Z, Love PED (2001) Outsourcing information systems: drawing lessons from a banking case study. Eur J Inf Syst 10(1):15–24

Bayramusta M, Nasir VA (2016) A fad or future of IT?: a comprehensive literature review on the cloud computing research. Int J Inf Manag 36(4):635–644

Benlian A, Hess T (2011) Opportunities and risks of software-as-a-service: findings from a survey of IT executives. Decis Support Syst 52(1):232–246

Benlian A, Kettinger WJ, Sunyaev A, Winkler TJ (2018) Special section: The transformative value of cloud computing: a decoupling, platformization, and recombination theoretical framework. J Manag Inf Syst 35(3):719–739

Bennett C, Timbrell GT (2000) Application service providers: will they succeed? Inf Syst Front 2 (2):195–211

Berman F, Hey T (2004) The scientific imperative. In: Foster I, Kesselman C (eds) The grid 2: blueprint for a new computing infrastructure. The Morgan Kaufmann series in computer architecture and design. Morgan Kaufman, San Francisco, CA, pp 13–24

Bharadwaj A, El Sawy OA, Pavlou PA, Venkatraman N (2013) Digital business strategy: toward a next generation of insights. MIS Q 37(2):471–482
Bhattacherjee A, Park SC (2014) Why end-users move to the cloud: a migration-theoretic analysis. Eur J Inf Syst 23(3):357–372
Böhm M, Leimeister S, Riedl C, Krcmar H (2011) Cloud computing – outsourcing 2.0 or a new business model for IT provisioning? In: Keuper F, Oecking C, Degenhardt A (eds) Application management: challenges – service creation – strategies. Gabler, Wiesbaden, pp 31–56
Bunker G, Thomson D (2006) Delivering utility computing: business-driver IT optimization. Wiley, Chichester
Burda D, Teuteberg F (2014) The role of trust and risk perceptions in cloud archiving — results from an empirical study. J High Technol Manag Res 25(2):172–187
Burnham TA, Frels JK, Mahajan V (2003) Consumer switching costs: a typology, antecedents, and consequences. J Acad Mark Sci 31(2):109
Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I (2009) Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Futur Gener Comput Syst 25(6):599–616
Cafaro M, Aloisio G (2011) Grids, clouds, and virtualization. In: Cafaro M, Aloisio G (eds) Grids, clouds and virtualization. Computer communications and networks. Springer, London, pp 1–21
Carr N (2008) The big switch: rewiring the world, from Edison to Google. W.W. Norton, New York, NY
Cloudscene (2018) Top ten data center operators in North America, EMEA, Oceania and Asia for the January to March 2018 period. https://cloudscene.com/top10. Accessed 11 Dec 2018
Cusumano M (2010) Cloud computing and SaaS as new computing platforms. Commun ACM 53 (4):27–29
Dašić P, Dašić J, Crvenković B (2016) Service models for cloud computing: search as a service (SaaS). Int J Eng Technol (IJET) 8(5):2366–2373
Dillon T, Wu C, Chang E (2010) Cloud computing: issues and challenges. Paper presented at the 24th IEEE international conference on advanced information networking and applications, Perth, WA, 20–23 Apr 2010
Doelitzscher F, Fischer C, Moskal D, Reich C, Knahl M, Clarke N (2012) Validating cloud infrastructure changes by cloud audits. Paper presented at the 8th IEEE world congress on services, Honolulu, HI, 24–29 June 2012
Durkee D (2010) Why cloud computing will never be free. Commun ACM 53(5):62–69
ENISA (2012) Cloud computing – benefits, risks and recommendations for information security. European network and security agency. https://resilience.enisa.europa.eu/cloud-security-and-resilience/publications/cloud-computing-benefits-risks-and-recommendations-for-information-security. Accessed 17 Sept 2019
Fernandes DAB, Soares LFB, Gomes JV, Freire MM, Inácio PRM (2014) Security issues in cloud environments: a survey. Int J Inf Secur 13(2):113–170
Foster I, Kesselman C (2004) The grid 2: blueprint for a new computing infrastructure, 2nd edn. Elsevier, San Francisco, CA
Foster I, Kesselman C, Tuecke S (2001) The anatomy of the grid: enabling scalable virtual organizations. Int J High Perform Comput Appl 15(3):200–222
Foster I, Zhao Y, Raicu I, Lu S (2008) Cloud computing and grid computing 360-degree compared. Paper presented at the grid computing environments workshop, Austin, TX, 12–16 Nov 2008
Gao F, Thiebes S, Sunyaev A (2018) Rethinking the meaning of cloud computing for health care: a taxonomic perspective and future research directions. J Med Internet Res 20(7):e10041
Gartner (2018a) Revenue of cloud computing worldwide. https://de.statista.com/statistik/daten/studie/195760/umfrage/umsatz-mit-cloud-computing-weltweit/. Accessed 9 Dec 2018
Gartner (2018b) Revenues from public cloud infrastructure as a service (IaaS) market worldwide from 2015 to 2017, by vendor (in million U.S. dollars). https://www.statista.com/statistics/754826/worldwide-public-cloud-infrastructure-services-vendor-revenues/. Accessed 11 Dec 2018
Gefen D, Karahanna E, Straub DW (2003) Trust and TAM in online shopping: an integrated model. MIS Q 27(1):51–90

Gonzalez R, Gasco J, Llop J (2009) Information systems outsourcing reasons and risks: an empirical study. Int J Soc Sci 4(3):180–191
Grover V, Kohli R (2012) Cocreating IT value: new capabilities and metrics for multifirm environments. MIS Q 36(1):225–232
Grozev N, Buyya R (2014) Inter-cloud architectures and application brokering: taxonomy and survey. Softw Pract Exp 44(3):369–390
Hentschel R, Leyh C, Petznick A (2018) Current cloud challenges in Germany: the perspective of cloud service providers. J Cloud Comput 7(1):5
Hess T, Matt C, Benlian A, Wiesböck F (2016) Options for formulating a digital transformation strategy. MIS Q Exec 15(2):123–139
Höfer CN, Karagiannis G (2011) Cloud computing services: taxonomy and comparison. J Internet Serv Appl 2(2):81–94
Hogendorn C (2011) Excessive(?) entry of national telecom networks, 1990–2001. Telecommun Policy 35(11):920–932
Hussain SA, Fatima M, Saeed A, Raza I, Shahzad RK (2017) Multilevel classification of security concerns in cloud computing. Appl Comput Inform 13(1):57–65
ISO (2004) Conformity assessment – vocabulary and general principles. https://www.iso.org/standard/29316.html. Accessed 17 Sept 2019
ITCandor (2018) Distribution of cloud platform as a service (PaaS) market revenues worldwide from 2015 to June 2018, by vendor. https://www.statista.com/statistics/540521/worldwide-cloud-platform-revenue-share-by-vendor/. Accessed 11 Dec 2018
Iyer B, Henderson JC (2010) Preparing for the future: understanding the seven capabilities of cloud computing. MIS Q Exec 9(2):117–131
Jones MA, Mothersbaugh DL, Beatty SE (2002) Why customers stay: measuring the underlying dimensions of services switching costs and managing their differential strategic outcomes. J Bus Res 55(6):441–450
Kern T, Lacity MC, Willcocks L (2002) Netsourcing: renting business applications and services over a network. Prentice-Hall, New York, NY
Khan N, Al-Yasiri A (2016) Identifying cloud security threats to strengthen cloud computing adoption framework. Procedia Comput Sci 94:485–490
Killalea T (2008) Meet the virts. ACM Queue 6(1):14–18
Kleinrock L (2005) A vision for the internet. ST J Res 2(1):4–5
Kouatli I (2014) A comparative study of the evolution of vulnerabilities in IT systems and its relation to the new concept of cloud computing. J Manag Hist 20(4):409–433
Kroeker KL (2011) Grid computing's future. Commun ACM 54(3):15–17
Krotsiani M, Spanoudakis G, Kloukinas C (2015) Monitoring-based certification of cloud service security. Paper presented at the OTM confederated international conferences "On the move to meaningful internet systems", Rhodes, 26–30 Oct 2015
Kumar N, DuPree L (2011) Protection and privacy of information assets in the cloud. In: Halpert B (ed) Auditing cloud computing, a security and privacy guide. Wiley, Hoboken, NJ, pp 97–128
Kunz I, Stephanow P (2017) A process model to support continuous certification of cloud services. Paper presented at the 31st IEEE international conference on advanced information networking and applications, Taipei, 27–29 Mar 2017
Lansing J, Benlian A, Sunyaev A (2018) 'Unblackboxing' decision makers' interpretations of IS certifications in the context of cloud service certifications. J Assoc Inf Syst 19(11):1064–1096
Lee C, Wan G (2010) Including subjective norm and technology trust in the technology acceptance model: a case of e-ticketing in China. SIGMIS Database 41(4):40–51
Leimeister S, Böhm M, Riedl C, Krcmar H (2010) The business perspective of cloud computing: actors, roles and value networks. Paper presented at the 18th European conference on information systems, Pretoria, 7–9 June 2010
Lins S, Grochol P, Schneider S, Sunyaev A (2016a) Dynamic certification of cloud services: trust, but verify! IEEE Secur Priv 14(2):66–71

Lins S, Teigeler H, Sunyaev A (2016b) Towards a bright future: enhancing diffusion of continuous cloud service auditing by third parties. Paper presented at the 24th European conference on information systems, Istanbul, 12–15 June 2016

Lins S, Schneider S, Sunyaev A (2018) Trust is good, control is better: creating secure clouds by continuous auditing. IEEE Trans Cloud Comput 6(3):890–903

Lins S, Schneider S, Sunyaev A (2019a) Cloud-Service-Zertifizierung: Ein Rahmenwerk und Kriterienkatalog zur Zertifizierung von Cloud-Services, 2nd edn. Springer, Berlin

Lins S, Schneider S, Szefer J, Ibraheem S, Sunyaev A (2019b) Designing monitoring systems for continuous certification of cloud services: deriving meta-requirements and design guidelines. Commun Assoc Inf Syst 44:460–510

Linthicum DS (2009) Cloud computing and SOA convergence in your enterprise: a step-by-step guide: how to use SaaS, SOA, Mashups, and web 2.0 to break down the IT gates. Addison-Wesley, Boston, MA

Liu F, Tong J, Mao J, Bohn R, Messina J, Badger L, Leaf D (2011) NIST cloud computing reference architecture. https://bigdatawg.nist.gov/_uploadfiles/M0008_v1_7256814129.pdf. Accessed 11 Dec 2018

Luftman J, Zadeh HS (2011) Key information technology and management issues 2010–11: an international study. J Inf Technol 26(3):193–204

Malluhi Q, Khan KM (2013) Trust in cloud services: providing more controls to clients. Computer 46(7):94–96

Marston S, Li Z, Bandyopadhyay S, Zhang J, Ghalsasi A (2011) Cloud computing — the business perspective. Decis Support Syst 51(1):176–189

Mell P, Grance T (2011) The NIST definition of cloud computing. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf. Accessed 17 Sept 2019

NIST Cloud Computing Security Working Group (2013) NIST cloud computing security reference architecture. https://csrc.nist.gov/publications/detail/sp/500-299/draft. Accessed 10 Dec 2018

Pahl C (2015) Containerization and the PaaS cloud. IEEE Cloud Comput 2(3):24–31

Park S-T, Park E-M, Seo J-H, Li G (2016) Factors affecting the continuous use of cloud service: focused on security risks. Clust Comput 19(1):485–495

Sarkar P, Young L (2011) Sailing the cloud: a case study of perceptions and changing roles in an Australian University. Paper presented at the 19th European conference on information systems, Helsinki, 9–11 June 2011

Schneider S, Sunyaev A (2016) Determinant factors of cloud-sourcing decisions: reflecting on the IT outsourcing literature in the era of cloud computing. J Inf Technol 31(1):1–31

Schneider S, Lansing J, Gao F, Sunyaev A (2014) A taxonomic perspective on certification schemes: development of a taxonomy for cloud service certification criteria. Paper presented at the 47th Hawaii international conference on system sciences, Waikoloa, HI, 6–9 Jan 2014

Schneider S, Wollersheim J, Krcmar H, Sunyaev A (2018) How do requirements evolve over time? a case study investigating the role of context and experiences in the evolution of enterprise software requirements. J Inf Technol 33(2):151–170

Schwarz A, Jayatilaka B, Hirschheim R, Goles T (2009) A conjoint approach to understanding IT application services outsourcing. J Assoc Inf Syst 10(10):1

Schwiegelshohn U, Badia RM, Bubak M, Danelutto M, Dustdar S, Gagliardi F, Geiger A, Hluchy L, Kranzlmüller D, Laure E, Priol T, Reinefeld A, Resch M, Reuter A, Rienhoff O, Rüter T, Sloot P, Talia D, Ullmann K, Yahyapour R (2010) Perspectives on grid computing. Futur Gener Comput Syst 26(8):1104–1115

Seethamraju R (2013) Determinants of SaaS ERP systems adoption. Paper presented at the 17th Pacific Asia conference on information systems, Jeju Island, 18–22 June 2013

Sehgal NK, Sohoni S, Xiong Y, Fritz D, Mulia W, Acken JM (2011) A cross section of the issues and research activities related to both information security and cloud computing. IETE Tech Rev 28(4):279–291

Sharma D, Dhote C, Potey MM (2016) Identity and access management as security-as-a-service from clouds. Procedia Comput Sci 79:170–174

Singh A, Chatterjee K (2017) Cloud security issues and challenges: a survey. J Netw Comput Appl 79(C):88–115
Singh S, Jeong Y-S, Park JH (2016) A survey on cloud computing security. J Netw Comput Appl 75(C):200–222
Soares J, Carapinha J, Melo M, Monteiro R, Sargento S (2011) Building virtual private clouds with network-aware cloud. Paper presented at the 5th international conference on advanced engineering computing and applications in sciences, Lisbon, 20–25 Nov 2011
Stephanow P, Banse C (2017) Evaluating the performance of continuous test-based cloud service certification. Paper presented at the 17th IEEE/ACM international symposium on cluster, cloud and grid computing, Madrid, 14–17 May 2017
Stephanow P, Khajehmoogahi K (2017) Towards continuous security certification of software-as-a-service applications using web application testing techniques. Paper presented at the 31st IEEE international conference on advanced information networking and applications, Taipei, 27–29 Mar 2017
Stephanow P, Banse C, Schütte J (2016) Generating threat profiles for cloud service certification systems. Paper presented at the 17th IEEE international symposium on high assurance systems engineering, Orlando, FL, 7–9 Jan 2016
Subashini S, Kavitha V (2011) A survey on security issues in service delivery models of cloud computing. J Netw Comput Appl 34(1):1–11
Sunyaev A, Schneider S (2013) Cloud services certification. Commun ACM 56(2):33–36
Susarla A, Barua A, Whinston AB (2003) Understanding the service component of application service provision: an empirical analysis of satisfaction with ASP services. MIS Q 27(1):91–123
Tan W, Fan Y, Ghoneim A, Hossain MA, Dustdar S (2016) From the service-oriented architecture to the web API economy. IEEE Internet Comput 20(4):64–68
Teigeler H, Lins S, Sunyaev A (2018) Drivers vs. inhibitors – what clinches continuous service certification adoption by cloud service providers? Paper presented at the 51th Hawaii international conference on system sciences, Hilton Waikoloa Village, HI, 3–6 Jan 2018
Thiebes S, Kleiber G, Sunyaev A (2017) Cancer genomics research in the cloud: a taxonomy of genome data sets. Paper presented at the 4th international workshop on genome privacy and security, Orlando, FL, 15 Oct 2017
Tiwana A, Konsynski B, Bush AA (2010) Research commentary—platform evolution: coevolution of platform architecture, governance, and environmental dynamics. Inf Syst Res 21(4):675–687
Top Threats Working Group (2016) The treacherous 12. Cloud computing top threats in 2016. https://downloads.cloudsecurityalliance.org/assets/research/top-threats/Treacherous-12_Cloud-Computing_Top-Threats.pdf. Accessed 17 Sept 2019
Trenz M, Huntgeburth JC, Veit DJ (2013) The role of uncertainty in cloud computing continuance: antecedents, mitigators, and consequences. Paper presented at the 21st European conference on information systems, Utrecht, 5–8 June 2013
Trenz M, Huntgeburth J, Veit D (2018) Uncertainty in cloud service relationships: uncovering the differential effect of three social influence processes on potential and current users. Inf Manag 55(8):971–983
Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27(3):425–478
Venters W, Whitley EA (2012) A critical review of cloud computing: researching desires and realities. J Inf Technol 27(3):179–197
Voorsluys W, Broberg J, Buyya R (2011) Introduction to cloud computing. In: Buyya R, Broberg J, Goscinski A (eds) Cloud computing. Wiley, Hoboken, NJ, pp 3–42
Wang B, Li B, Li H (2014) Oruta: privacy-preserving public auditing for shared data in the cloud. IEEE Trans Cloud Comput 2(1):43–56
Weinhardt C, Anandasivam A, Blau B, Borissov N, Meinl T, Michalk W, Stößer J (2009) Cloud computing–a classification, business models, and research directions. Bus Inf Syst Eng 1 (5):391–399
Whitman M, Mattord H (2011) Cloud-based IT governance. In: Halpert B (ed) Auditing cloud computing: a security and privacy guide. Wiley, Hoboken, NJ, pp 33–55

Windhorst I, Sunyaev A (2013) Dynamic certification of cloud services. Paper presented at the 8th international conference on availability, reliability and security, Regensburg, 2–6 Sept 2013
Yoo Y, Henfridsson O, Lyytinen K (2010) Research commentary—the new organizing logic of digital innovation: an agenda for information systems research. Inf Syst Res 21(4):724–735

Further Reading

Benlian A, Kettinger WJ, Sunyaev A, Winkler TJ (2018) Special section: The transformative value of cloud computing: a decoupling, platformization, and recombination theoretical framework. J Manag Inf Syst 35(3):719–739
Fernandes DAB, Soares LFB, Gomes JV, Freire MM, Inácio PRM (2014) Security issues in cloud environments: a survey. Int J Inf Secur 13(2):113–170
Lins S, Schneider S, Sunyaev A (2019) Cloud-Service-Zertifizierung: Ein Rahmenwerk und Kriterienkatalog zur Zertifizierung von Cloud-Services, 2nd edn. Springer, Berlin
Marston S, Li Z, Bandyopadhyay S, Zhang J, Ghalsasi A (2011) Cloud computing — the business perspective. Decis Support Syst 51(1):176–189