CHAPTER

Introduction 1

Computer clouds are utilities providing computing services. In *utility computing*, the hardware and the software resources are concentrated in large data centers, and users of computing services pay as they consume computing, storage, and communication resources. While utility computing often requires a cloud-like infrastructure, the focus of cloud computing is on the business model for providing computing services.

NIST, the US National Institute of Standards and Technology, defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction." Cloud computing is characterized by five attributes: *on-demand self-service, broad network access, resource pooling, rapid elasticity*, and *measured service*.

More than half a century ago, at the centennial anniversary of MIT, John McCarthy, the 1971 Turing Award recipient for his work in artificial intelligence, prophetically stated: "... If computers of the type I have advocated become the computers of the future, then computing may someday be organized as a public utility, just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry." McCarthy's prediction is now a technological and social reality.

The cloud computing movement is motivated by the idea that data processing and storage can be done more efficiently on large farms of computing and storage systems accessible via the Internet. Computer clouds support a paradigm shift from local to network-centric computing and network-centric content where distant data centers provide the computing and storage resources. In this new paradigm, users relinquish control of their data and code to Cloud Service Providers (CSPs).

The cloud computing age started in 2006 when Elastic Cloud Computing (EC2) and Simple Storage Service (S3) were offered by Amazon Web Services (AWS). Six years later, in 2012, EC2 was used by businesses in 200 countries. S3 has surpassed two trillion objects, and routinely runs more than 1.1 million peak requests per second. The range of services offered by CSPs and the number of cloud users have increased dramatically during the last few years. Cloud computing offers scalable and elastic computing and storage services. The resources used for these services can be metered, and the users can be charged only for the resources they have used. Cloud computing is a business reality, as a large number of organizations have adopted this paradigm.

Internet users have discovered the appeal of cloud computing either directly or indirectly, through a variety of services, without knowing the role the clouds play in their lives. The number of cloud users will continually increase in the years to come as the vast computational resources provided by the cloud infrastructure and the exabytes of data stored in the clouds are streamed, downloaded, accessed and used for deep learning, designing and engineering complex systems, scientific discovery, education, business, analytics, art, and virtually all other aspects of human endeavor.

Cloud Computing. https://doi.org/10.1016/B978-0-32-385277-7.00008-7

Data analytics, data mining, computational financing, scientific and engineering applications, gaming, and social networking, as well as other computational and data-intensive activities benefit from cloud computing. Content previously confined to personal devices, such as workstations, laptops, tablets, and smartphones, no longer need to be stored locally. Data stored on computer clouds can be shared among all these devices, and it is accessible whenever a device is connected to the Internet. Clouds continually evolve in predictable, as well as unpredictable ways; for example, edge computing proposes to minimize network traffic and response time by preprocessing data locally.

Almost half a century after the dawn of the computing era, an eternity in the age of the silicon, disruptive multicore technology and the enormous computing power of Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs) challenge computer science community to develop new algorithms, programming environments, and tools and challenge application developers to exploit concurrency. There is no point now to wait for faster clock rates; we better design algorithms and applications able to use modern processors and co-processors. The new challenge is harnessing the power of millions of multicore processors and systems on a chip and allowing them to work in concert efficiently.

Cloud computing has shown that there are applications that can effortlessly exploit concurrency and, in the process, generate huge revenues. A new era in computing has begun, a time when Big Data hides nuggets of useful information and requires massive amounts of computing resources. In this era, "coarse" is good and "fine" is not good, at least as far as the granularity of parallelism is concerned.

The scale of computer clouds amplifies unanticipated benefits, as well as the nightmares of system designers. Even a slight improvement of server performance and/or of the algorithms for resource management could lead to huge cost savings and rave reviews. When engineering large-scale systems, an important lesson is to prepare for the unexpected because low probability events occur and can cause major disruptions. The failure of one of the millions of hardware and software components can be amplified, propagate throughout the entire system, and have catastrophic consequences.

Cloud computing is a *disruptive computing paradigm* and requires major changes in many areas of computer science and computer engineering, including data storage, computer architecture, networking, resource management, scheduling, and last but not least computer security. Computer clouds operate in an environment characterized by the *variability of everything* and by *conflicting requirements*. Such disruptive qualities of computer clouds ultimately demand new thinking in system design. This book covers challenges posed by the scale of the cloud infrastructure and the large population of cloud users with diverse applications and requirements.

Cloud computing is cost-effective because of *resource multiplexing.* Application data is stored closer to the site where it is used in a manner that is device and location-independent; potentially, this data-storage strategy increases reliability, as well as security. Organizations using computer clouds are relieved of supporting large IT teams, acquiring and maintaining costly hardware and software, and paying large electricity bills. CSPs can operate more efficiently due to economies of scale.

Cloud computing represents a dramatic shift in the design of systems capable of providing vast amounts of computing cycles and storage space. *Computer clouds use off-the-shelf, low-cost components.* During the previous four decades, powerful, one-of-a-kind systems were built at a high cost, with the most advanced components available at the time. In early 1990s, Gordon Bell argued that one-of-a-kind systems are not only expensive to build but also that the cost of rewriting applications for them is prohibitive. He anticipated that sooner or later massively parallel computing will evolve into *computing for the masses* [55].

Since there are virtually no bounds on the composition of digital systems controlled by software, we are tempted to build increasingly more complex systems, including systems of systems [340]. The behavior and the properties of such systems are not always well understood thus, we should not be surprised that large-scale systems will occasionally fail and computing clouds will occasionally exhibit an unexpected behavior.

Cloud computing reinforces the idea that *computing and communication are deeply intertwined.* Advances in one field are also critical for the other. Cloud computing would not have emerged as an alternative to traditional computing models before the Internet was able to support high-bandwidth, reliable, low-cost communication. High-performance switches are critical elements of supercomputers and clouds infrastructure. Internet routers use powerful processors and Artificial Intelligence (AI) to enhance security.

The architecture, the coordination mechanisms, the design methodology, and the analysis techniques for large-scale complex systems such as clouds will evolve in response to changes in technology, the environment, and the demands of user community. Some of these changes will reflect changes in communication and in the Internet itself in terms of speed, reliability, security, capacity to accommodate a larger addressing space by migration to IPv6, and so on. The complexity of the cloud computing infrastructure is unquestionable and raises questions such as: How can we manage such systems? Do we have to consider radically new ideas, such as self-management and self-repair for future clouds consisting of millions of servers? Should we migrate from a strictly deterministic view of such complex systems to a non-deterministic one? Answers to these questions provide a rich set of research topics for the computer science and engineering community.

The cloud movement is not without skeptics and critics. The critics argue that cloud computing is just a marketing ploy, that users may become dependent on proprietary systems, and that the failure of a large system such as the cloud could have significant consequences for a very large group of users who depend on the cloud for their computing and storage needs. Security and privacy are major concerns for cloud computing users.

1.1 Cloud computing, an old idea whose time has come

It is hard to pinpoint a single technological or architectural development that triggered the movement towards computer clouds. This movement is the result of a cumulative effect of developments in microprocessors, storage, and networking technologies coupled with architectural advancements in all these areas and with advances in software systems, tools, programming languages, and algorithms supporting distributed and parallel computing.

Through the years, we have witnessed the breathtaking evolution of solid-state technologies that led to the development of multicore processors. The proximity of multiple cores on the same silicon die allows cache-coherency circuitry to operate at a much higher clock rate than would be possible if signals were to travel off-chip. Systems on a chip (SoCs), like Apple's M1, with general-purpose cores, GPU cores, and TPU cores, deliver impressive computing power with lower power consumption.

Storage technology has also evolved dramatically. Solid-state disks enable systems to manage very high transaction volumes and larger numbers of concurrent users while the price of memory has dropped significantly. Optical storage technologies and flash memories are widely used nowadays.

The thinking in software engineering has also evolved, and new models have emerged. A software architecture and a software design pattern, namely, the *three-tier model,* has emerged. Its components are discussed next. The *presentation tier* is the topmost level of the application; typically, it runs on a desktop or laptop, uses a standard graphical user interface, and displays information related to services. The *application/logic tier* controls the functionality of an application, may consist of one or more separate modules running on a workstation or application server, and may be multi-tiered itself. The *data tier* controls the servers where the information is stored; it runs a relational database management system on a database server or a mainframe and contains the computer data storage logic; it keeps data independent from application servers or processing logic and improves scalability and performance. Any of the tiers can be replaced independently: for example, a change of operating system in the presentation tier would only affect the user interface code.

Once the technological elements were in place, it was only a matter of time until the economical advantages of cloud computing became apparent. Due to the economy of large-scale data centers, centers with more than 50 000 systems are more economical to operate than medium-sized centers that have around 1 000 systems. Large data centers equipped with commodity computers experience a five to seven times decrease in resource consumption, including energy, compared to medium-sized data centers [32].

Several factors contribute to the success of cloud computing: (i) technological advances; (ii) a realistic system model; (iii) user convenience, and (iv) cost. A non-exhaustive list of cloud computing advantages over previous attempts to network centric computing includes:

- Cloud computing is in a better position to exploit recent advances in software, networking, storage, and processor technologies. Cloud computing is promoted by large IT companies where technological developments take place; the companies have a vested interest to promote the new technologies.
- A cloud consists of hardware and software resources in a single administrative domain where resource management, fault-tolerance, and quality of service are less challenging than distributed computing with resources in multiple administrative domains.
- Cloud computing is focused on enterprise computing [159,164]; its adoption by industrial organizations, financial institutions, healthcare organizations, and so on, has a potentially huge economic impact.
- A cloud provides the illusion of infinite computing resources; computer cloud elasticity frees applications designers from the confinements of a single system.
- A cloud eliminates the need for up-front financial commitment, and it is based on a pay-as-you-go approach; this has the potential to attract new applications and new users for existing applications, fomenting a new era of industry-wide technological advancements.

The term "computer cloud" covers infrastructures of different sizes, with different management and different user populations. Several types of clouds can be identified: (i) *public cloud*—the infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services; (ii) private cloud—the infrastructure is operated solely for an organization; (iii) hybrid cloud—the infrastructure is a composition of two or more clouds (e.g., public and private) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability; and (iv) community cloud—the infrastructure is shared by several organizations and supports a specific community with shared concerns.

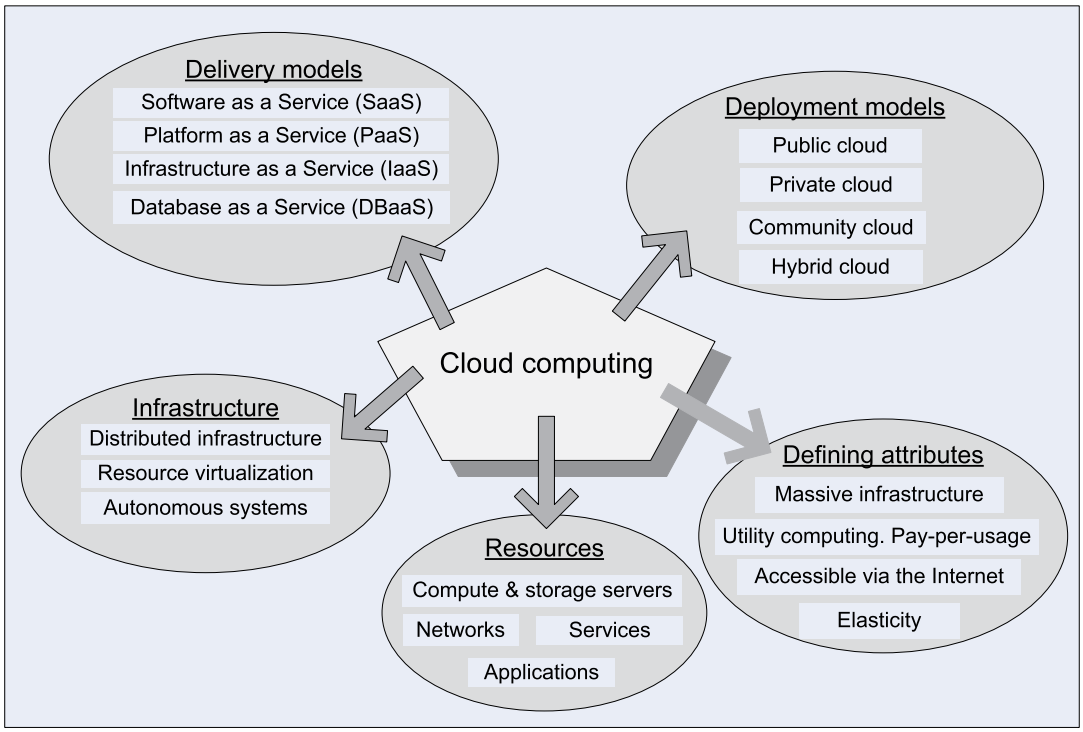

FIGURE 1.1

A summary view of cloud computing: delivery models, deployment models, defining attributes, and organization of cloud infrastructure, discussed in Chapter 2, and cloud resources, discussed in Chapters 4, 6, 7.

A private cloud could provide the computing resources needed for a large organization, e.g., a research institution, a university, or a corporation. The argument that a private cloud does not support utility computing is based on the observation that an organization has to invest in the infrastructure and a user of a private cloud does pay as it consumes resources [32]. Nevertheless, a private cloud could use the same hardware infrastructure as a public one; its security requirements will be different from those for a public cloud, and the software running on the cloud is likely to be restricted to a specific domain.

Cloud computing delivery models, deployment models, defining attributes, resources, and organizations of the infrastructure are summarized in Fig. 1.1. The defining attributes of the new philosophy for delivering computing services are:

- Cloud computing uses Internet technologies to offer elastic services, i.e., the ability of dynamically acquiring computing resources and supporting a variable workload. A cloud service provider maintains a massive infrastructure to support elastic services.
- Resources used for these services can be metered, and users can be charged only for the resources they used.
- Maintenance and security are ensured by service providers.

- Economy of scale allows service providers to operate more efficiently due to specialization and centralization.
- Cloud computing is cost-effective due to resource multiplexing; lower costs for the service provider are passed on to cloud users.
- The application data is stored closer to the site where it is used in a device- and location-independent manner; potentially, this data storage strategy increases reliability and security and, at the same time, lowers communication costs.

In spite of the technological breakthroughs that have made cloud computing feasible, there are still major obstacles for this new technology; these obstacles provide opportunities for research. We list a few of the most obvious obstacles:

- Availability of service; what happens when the service provider cannot deliver? Can a large company such as GM move its IT to the cloud and have assurances that its activity will not be negatively affected by cloud overload? A partial answer to this question is provided by Service Level Agreements (SLA)s discussed in Section 2.8.
- Performance unpredictability, unavoidable due to resource sharing.
- Elasticity, the ability to scale up and down quickly and *overprovisioning*, maintaining pools of resources considerably larger than the average need.
- Vendor lock-in; once a customer is hooked to one provider, it is hard to move to another. The standardization efforts at NIST attempt to address this problem.
- Data confidentiality and auditability, a serious concern analyzed in Chapter 8.
- Data transfer bottlenecks critical for data-intensive applications. Transferring 1 TB of data on a 1 Mbps network takes 8 000 000 seconds or about 10 days; it is faster and cheaper to use courier service and send data recoded on some media than to send it over the network. High-speed networks will alleviate this problem, e.g., a 1 Gbps network would reduce this time to 8 000 seconds, or slightly more than two hours.

Cloud computing is a technical and social reality and an emerging technology. At this time, one can only speculate how the infrastructure for this new paradigm will evolve and what applications will migrate to it. The economical, social, ethical, and legal implications of this shift in technology, when the users rely on services provided by large data centers and store private data and software on systems they do not control, are likely to be significant.

Scientific and engineering applications, deep learning, data mining, computational financing, and gaming and social networking, as well as many other computational and data-intensive activities, can benefit from cloud computing. A broad range of data from the results of high-energy physics experiments to financial or enterprise management data, to personal data such as photos, videos, and movies, can be stored on the cloud.

The obvious advantage of network-centric content is the accessibility of information from any site where one can connect to the Internet. Clearly, information stored on a cloud can be shared easily, but this approach also raises major concerns: Is the information safe and secure? Is it accessible when we need it? Do we still own it?

In the near future, the focus of cloud computing is expected to shift from building the infrastructure, today's main front of competition among the vendors, to the application domain. This shift in focus is reflected by Google's strategy to build a dedicated cloud for government organizations in the United

States. The Internet made cloud computing possible: We could not even dream of using computing and storage resources from distant data centers without fast communication. The evolution of cloud computing is organically tied to the future of the Internet. The Internet of Things (IoT) has already planted some of its early seeds in computer clouds, and Amazon already offers services such as Lambda and Kinesis.

In a discussion of the technology trends, Jim Gray emphasized that the cost of communication in a wide area network has decreased dramatically and will continue to do so. Thus, it makes economical sense to store the data near the application [207], in other words, to store it in the cloud where the application runs. This insight leads us to believe that several new classes of cloud computing applications could emerge in the next few years [32].

The excitement due to cloud computing has translated into a flurry of publications. In this book, we attempt to sift through the large volume of information and dissect the main ideas related to cloud computing. We first discuss applications of cloud computing and then analyze the infrastructure for cloud computing. Several decades of research in parallel and distributed computing have paved the way for cloud computing. Through the years, we have discovered the challenges posed by the implementation of parallel and distributed systems and the ways to address some of them while avoiding the others.

1.2 Energy use and ecological impact of cloud computing

The discussion of cloud infrastructure cannot be concluded without an analysis of the energy used for cloud computing and its impact on the environment. The energy consumption required by different types of human activities is partially responsible for greenhouse-gas emissions and has grave implications on climate change. Fig. 1.2 shows that industrialized countries, including the US and China, contribute significantly to CO_2 emissions according to data published in August 2020 by the Union of Concerned Scientists (see https://www.ucsusa.org/resources/each-countrys-share-co2-emissions). The data centers and the IT industries contribute to the large CO_2 footprint of these countries.

Reduction of energy consumption and thus of the carbon footprint of cloud-related activities is increasingly more important for society. Indeed, more and more applications run on clouds, and cloud computing uses more energy than many other human-related activities. Reduction of the carbon footprint can only be achieved through a comprehensive set of technical efforts. The hardware of the cloud infrastructure has to be refreshed periodically, and new and more energy efficient technologies have to be adopted; the resource management software has to pay more attention to energy optimization.

International Data Corporation (IDC) forecasts that continued adoption of cloud computing could prevent the emission of more than 1 billion metric tons of carbon dioxide (CO_2) from 2021 through 2024. This projection is based on the assumption that 60% of data centers will adopt the technology and processes underlying more sustainable "smarter" data centers by 2024 (https://www.environmentalleader.com/2021/03/report-continued-adoption-of-cloud-computing-could-prevent-emission-of-1-billion-tons-of-co2/). Several factors could contributes to lowering CO_2 emissions by cloud data centers: (i) shifting to cleaner sources of energy; (ii) reducing wasted energy and having more energy spent on running the IT equipment rather than on cooling; and (iii) delivering IT service wherever needed, by shifting workloads to enable greater use of renewable resources.

A 2020 paper published by the journal *Science* [336] supports this optimistic view in spite of the significant increase of data-center activity: "By 2018, global data center workloads and compute in-

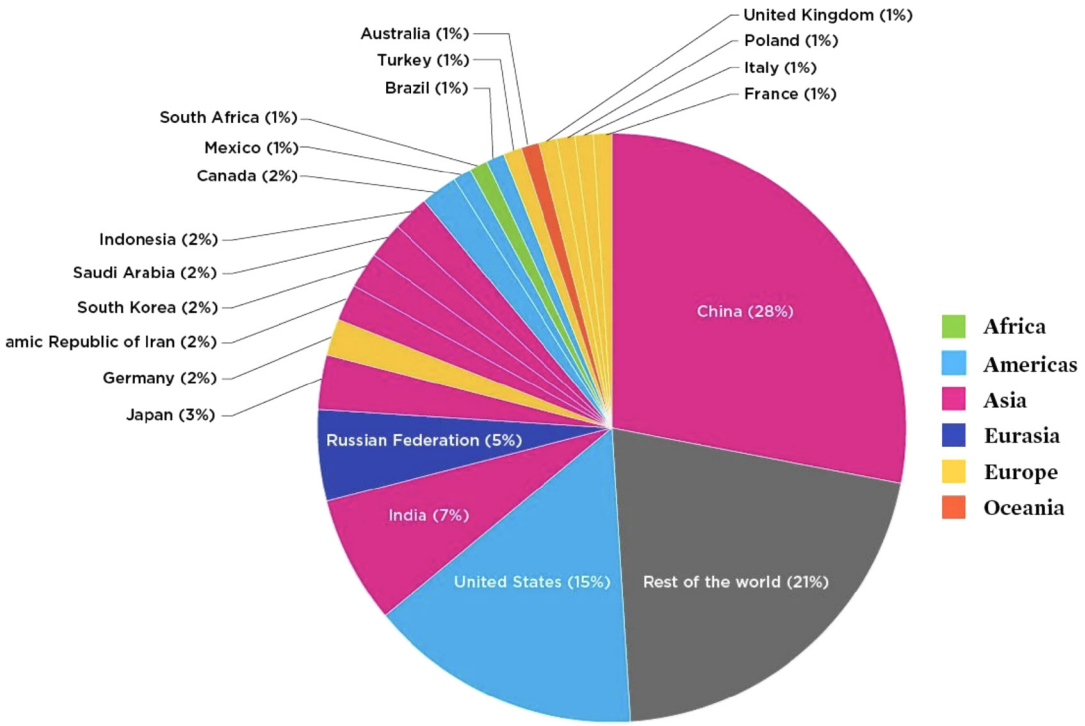

FIGURE 1.2

The CO_2 contribution of various countries of the world.

stances had increased more than sixfold, whereas data center internet protocol (IP) traffic had increased by more than 10-fold. Data center storage capacity has also grown rapidly, increasing by an estimated factor of 25 over the same time period... In 2018, we estimated that global data center energy use rose to 205 TWh, or around 1% of global electricity consumption. This represents a 6% increase compared with 2010, whereas global data center compute instances increased by 550% over the same time period. Expressed as energy use per compute instance, the energy intensity of global data centers has decreased by 20% annually since 2010, a notable improvement."

The website of the Center of Expertise for Energy Efficiency in Data Centers at Lawrence Berkeley National Laboratory, https://datacenters.lbl.gov/resources/recalibrating-global-data-center-energy, provides a wealth of information, tools, technology, and training for data-center energy optimization.

1.3 Ethical issues in cloud computing

Cloud computing is based on a paradigm shift with profound implications for computing ethics. The main elements of this shift are: (i) Control is relinquished to third party services; (ii) data is stored on

multiple sites administered by several organizations; and (iii) multiple services interoperate across the network. The complex structure of cloud services makes it difficult to determine who is responsible for each realm of action. Many entities contribute to an action with undesirable consequences, and no one can be held responsible. As a result of de-perimeterization, "not only the border of the organizations IT infrastructure blurs, also the border of the accountability becomes less clear" [477].

Unauthorized access, data corruption, infrastructure failure, and service unavailability are some of the risks related to relinquishing control to third-party services; moreover, whenever a problem occurs, it is difficult to identify the source and the entity causing it. Systems can span the boundaries of multiple organizations and cross the security borders, a process called *de-perimeterization*.

Ubiquitous and unlimited data sharing and storage among organizations test the self-determination of information, the right or ability of individuals to exercise personal control over the collection, and the use and the disclosure of their personal data by others. This tests the confidence and trust in today's evolving information society. Identity fraud and theft are made possible by unauthorized access to personal data in circulation and by new forms of dissemination through social networks. All these factors can also pose a danger to cloud computing.

Cloud service providers have already collected petabytes of sensitive personal information stored by data centers around the world. The acceptance of cloud computing will be determined by the effort dedicated by the CSPs and the countries where the data centers are located to ensure privacy. Privacy is affected by cultural differences: While some cultures favor privacy, other cultures emphasize community, and this leads to an ambivalent attitude towards privacy on the Internet, which is a global system.

The question of what can be done proactively about the ethics of cloud computing does not have easy answers because many undesirable phenomena in cloud computing will only become apparent over time. However, the need for rules and regulations for the governance of cloud computing are obvious. Governance means the manner in which something is governed or regulated, the method of management, and the system of regulations. Explicit attention to ethics must be paid by governmental organizations providing research funding; private companies are less constrained by ethics oversight, and governance arrangements are more conducive to profit generation.

Accountability is a necessary ingredient of cloud computing; adequate information about how data is handled within the cloud and about allocation of responsibility are key elements for enforcing ethics rules in cloud computing. Recorded evidence enables us to assign responsibility, but there can be tension between privacy and accountability, and it is important to establish what is being recorded and who has access to the records. Unwanted dependency on a cloud service provider, the so-called *vendor lock-in*, is a serious concern, and the current standardization efforts at NIST attempt to address this problem. Another concern for the users is a future in which only a handful of companies dominate the market and dictate prices and policies.

1.4 Factors affecting cloud service availability

Clouds are affected by malicious attacks and failures of the infrastructure, e.g., power failures. Such events can affect the Internet domain name servers and prevent access to a cloud or can directly affect the clouds. For example, an attack at Akamai on June 15, 2004, caused a domain name outage and a major blackout that affected Google, Yahoo, and many other sites. In May 2009, Google was the target

of a serious denial of service (DNS) attack that took down services like Google News, and Gmail for several days.

Lightning caused a prolonged down time at Amazon on June 29–30, 2012; the AWS cloud in the East region of the US, which consists of ten data centers across four availability zones, was initially troubled by utility power fluctuations, probably caused by an electrical storm. Availability zones are locations within data-center regions where public cloud services originate and operate. A June 29, 2012, storm on the East Coast took down some of Virginia-based Amazon facilities and affected companies using systems exclusively in this region. Instagram, a photo sharing service, was one of the victims of this outage.

The recovery from the failure took a very long time and exposed a range of problems. For example, one of the ten centers failed to switch to backup generators before exhausting the power that could be supplied by UPS units. AWS uses "control planes" to enable users to switch to resources in a different region, and this software component also failed. The booting process was faulty and extended the time to restart EC2 and EBS services. Another critical problem was a bug in the Elastic Load Balancer (ELB), which is used to route traffic to servers with available capacity. A similar bug affected the recovery process of Relational Database Service (RDS). This event brought to light "hidden" problems that occur only under specific circumstances.

The stability risks due to interacting services are discussed in [182]. A cloud application provider, a cloud storage provider, and a network provider could implement different policies, and the unpredictable interactions between load-balancing and other reactive mechanisms could lead to dynamic instabilities. The unintended coupling of independent controllers that manage the load, the power consumption, and the elements of the infrastructure could lead to undesirable feedback and instability similar to the one experienced by the policy-based routing in the Internet BGP (Border Gateway Protocol).

For example, the load balancer of an application provider could interact with the power optimizer of the infrastructure provider. Some of these couplings may only become manifest under extreme conditions and be very hard to detect under normal operating condition but could have disastrous consequences when the system attempts to recover from a hard failure, as in the case of the AWS 2012 failure.

Clustering resources in data centers located in different geographical areas lowers the probability of catastrophic failures. This geographic dispersion of resources could have additional positive side effects, such as reduction of communication traffic, lowering energy costs by dispatching the computations to sites where the electric energy is cheaper, and improving performance by an intelligent and efficient load-balancing strategy.

1.5 Network-centric computing and network-centric content

Network-centric computing and network-centric content imply that data processing and data storage can be done on computers accessed through the Internet. The term *content* refers to any type or volume of media, be it static or dynamic, monolithic or modular, live or stored, produced by aggregation or mixed. *Information* is the result of functions applied to content.

The content should be treated as having meaningful semantic connotations rather than a string of bytes; the focus will be on the information that can be extracted by content mining when users request

named data and content providers publish data objects. Content-centric routing will enable users to fetch the desired data from the most suitable location in terms of network latency or download time. In turn, the creation and consumption of audio and visual content is likely to transform the Internet to support increased quality in terms of resolution, frame rate, color depth, stereoscopic information, etc.

There are also some challenges, such as providing secure services for content manipulation, ensuring global rights-management, control over unsuitable content, and reputation management. Network-centric computing and network-centric contents share a number of characteristics:

1. Most applications are data intensive. Data analytics enable enterprises to optimize their operations; computer simulation is a powerful tool for scientific research in virtually all areas of science, from physics, biology, and chemistry to archeology. The widespread use of sensors contributes to the increase of the volume of data. Artificial Intelligence (AI) and Machine Learning (ML) algorithms require massive amounts of data.
2. Computing and communication resources (CPU cycles, storage, network bandwidth) are shared, and resources can be aggregated to support data-intensive applications. Multiplexing leads to a higher resource utilization: When multiple applications share a system, their peak demands for resources are not synchronized, and average system utilization increases.
3. Data sharing facilitates collaborative activities. Indeed, many applications in science and engineering, as well as industrial, financial, and governmental applications, require multiple types of analysis of shared-data sets and multiple decisions carried out by groups scattered around the globe. Open software development sites are another example of such collaborative activities.
4. Virtually all applications are network intensive. Indeed, transferring large volumes of data requires high bandwidth networks. Parallel computing, computation steering, and data streaming are examples of applications that can only run efficiently on low-latency networks. Computation steering in numerical simulation means to interactively guide a computational experiment towards a region of interest.
5. The systems are accessed using *thin clients* running on systems with limited resources.

There are sources of concern regarding the paradigm shift from locally owned resources to network-centric computing: (i) the management of large pools of resources poses new challenges because such systems are vulnerable to malicious attacks that can affect a large population of users; (ii) large-scale systems are affected by phenomena characteristic to complex systems such as phase transitions when a relatively small change in the environment can lead to an undesirable system state [332]; (iii) ensuring Quality of Service (QoS) guarantees is extremely challenging because perfect performance isolation and the ability to accommodate workloads with very large peak-to-average ratios are elusive; and (iv) data sharing poses not only security and privacy challenges but also requires mechanisms for access control for authorized users and for detailed logs of the history of data changes.