4 Cloud Basics – An Introduction to Cloud Computing

Katarina Stanoevska-Slabeva, Thomas Wozniak

4.1 Introduction

Cloud Computing has attracted a lot of attention in recent times. The media as well as analysts are generally very positive about the opportunities Cloud Computing is offering. In May 2008, Merrill Lynch (2008) estimated the cost advantages of Cloud Computing to be three to five times for business applications and more than five times for consumer applications. According to a Gartner press release from June 2008, Cloud Computing will be "no less influential than e-business" (Gartner 2008a).

The positive attitude towards the importance and influence of Cloud Computing resulted in optimistic Cloud-related market forecasts. In October 2008, IDC (2008b) forecasted an almost threefold growth of spending on Cloud services until 2012, reaching $42 billion. Same analyst firm reported that the cost advantage associated with the Cloud model becomes even more attractive in the economic downturn (IDC 2008b). Positive market prospects are also driven by the expectation that Cloud Computing might become the fundamental approach towards Green IT.

Despite of the broad coverage of Cloud Computing in commercial press, there is still no common agreement on what exactly Cloud Computing is and how it relates to Grid Computing. To gain an understanding of what Cloud Computing is, we first look at several existing definitions of the term. Based on those definitions, we identify key characteristics of Cloud Computing. Then we describe the common architecture and components of Clouds in detail, discuss opportunities and challenges of Cloud Computing, and provide a classification of Clouds. Finally, we make a comparison between Grid Computing and Cloud Computing.

4.2 Cloud Definitions

The term Cloud Computing has been defined in many ways by analyst firms, academics, industry practitioners, and IT companies. Table 4.1 shows how selected analyst firms define or describe Cloud Computing.

K. Stanoevska-Slabeva et al. (eds.), *Grid and Cloud Computing: A Business Perspective on Technology and Applications*, DOI 10.1007/978-3-642-05193-7_4,

Table 4.1: Cloud Computing definitions by selected analyst firms

Source	Definition
Gartner	"a style of computing in which massively scalable IT-related capabilities are provided "as a service" using Internet technologies to multiple external customers" (Gartner 2008b)
IDC	"an emerging IT development, deployment and delivery model, enabling real-time delivery of products, services and solutions over the Internet (i.e., enabling cloud services)" (Gens 2008)
The 451 Group	"a service model that combines a general organizing principle for IT delivery, infrastructure components, an architectural approach and an economic model – basically, a confluence of grid computing, virtualization, utility computing, hosting and software as a service (SaaS)" (Fellows 2008)
Merrill Lynch	"the idea of delivering personal (e.g., email, word processing, presentations.) and business productivity applications (e.g., sales force automation, customer service, accounting) from centralized servers" (Merrill Lynch 2008)

All these definitions have a common characteristic: they try to describe and define Cloud Computing from the perspective of the end users and their focus is on how it might be experienced by them. According to these definitions, core feature of Cloud Computing is the provision of IT infrastructure and applications as a service in a scalable way.

The definition of Cloud Computing has been subject of debate also in the scientific community. Similar to the commercial press, there are different opinions about what Cloud Computing is and which features distinguish a Cloud. Compared to the definitions from the commercial press, the definitions in scientific literature include not only the end user perspective, but also architectural aspects. For example, Berkeley RAD Lab define Cloud Computing as follows:

> "Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what we will call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, we call it a Public Cloud; the service being sold is Utility Computing. We use the term Private Cloud to refer to internal datacenters of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds. People can be users or providers of SaaS, or users or providers of Utility Computing." (Armbrust et al. 2009)

This definition unites different perspectives on a Cloud: from the perspective of a provider, the major Cloud component is the data centre. The data centre contains the raw hardware resources for computing and storage, which together with software are offered in a pay-as-you-go manner. From the perspective of their purpose, Clouds are classified into private and public. Independent of the purpose of Clouds, one most important characteristic of Clouds is the integration of hardware and system software with applications, i.e. integration of utility computing and SaaS.

Also Reese (2009) notes a Cloud can be both software and infrastructure, and stresses the way how Cloud services might be consumed:

> "The [Cloud] service is accessible via a web browser (nonproprietary) or web services API.; Zero capital expenditure is necessary to get started.; You pay only for what you use as you use it."

Foster et al. (2008) define Cloud Computing as

> "[a] large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet."

Two important aspects added by the definition of Foster et al. (2008) are virtualization and scalability. Cloud Computing abstracts from the underlying hardware and system software through virtualization. The virtualized resources are provided through a defined abstracting interface (an Application Programming Interface (API) or a service). Thus, at the raw hardware level, resources can be added or withdrawn according to demand posted through the interface, while the interface to the user is not changing. This architecture enables scalability and flexibility on the physical layer of a Cloud without impact on the interface to the end user.

Finally, Vaquero et al. (2008) analysed no less than 22 definitions of Cloud Computing, all proposed in 2008. Based on that analysis, Vaquero et al. (2008) propose the following definition which aims to reflect how Cloud Computing is currently conceived:

> "Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs."

Further, Vaquero et al. (2008) summarized *scalability*, *pay-per-use utility model* and *virtualization* as the feature set that would most closely resemble a minimum definition of Clouds. However, while the definition of Vaquero et al. (2008) summarizes other definitions with respect to the physical layer very well, it does not stress the integration of hardware with Software-as-a-Service in sufficient manner.

All definitions illustrate that Cloud Computing is a phenomenon that comprises a number of aspects and is related to a new paradigm of IT (hardware and applications) delivery and deployment. Generally, Cloud Computing concerns the delivery of IT capabilities to external customers, or, from the perspective of a user, obtaining IT capabilities from an external provider, as a service in a pay-per-use manner and over the Internet. Further, scalability and virtualization are very often seen as key characteristics of Cloud Computing (e.g. Foster et al. 2008, Sun 2009a, Vaquero et al. 2009). Scalability refers to a dynamic adjustment of provisioned IT resources to variable load, e.g. increasing or decreasing number of users, required storage capacity or processing power. Virtualization, which is also regarded as the cornerstone technology for all Cloud architectures (e.g. Sun 2009), is mainly used for abstraction and encapsulation (Foster et al. 2008). Abstraction allows unifying

raw compute, storage, and network resources as a pool of resources and building resource overlays such as data storage services on top of them (Foster et al. 2008). Encapsulation of applications ultimately improves security, manageability, and isolation (Foster et al. 2008). Another important feature of Clouds is the integration of hardware and system software with applications. Both the hardware and systems software, or infrastructure, and the applications are offered as a service in an integrated manner.

Based on the findings of the definition analysis, a summary of the defining features of Cloud Computing, as they will be applied to guide further discussions in this book, is provided below and in figure 4.1:

- Cloud Computing is a new computing paradigm.
- Infrastructure resources (hardware, storage and system software) and applications are provided in a X-as-a-Service manner. When these services are offered by an independent provider or to external customers, Cloud Computing is based on pay-per-use business models.
- Main features of Clouds are virtualization and dynamic scalability on demand.
- Utility computing and SaaS are provided in an integrated manner, even though utility computing might be consumed separately.
- Cloud services are consumed either via Web browser or via a defined API.

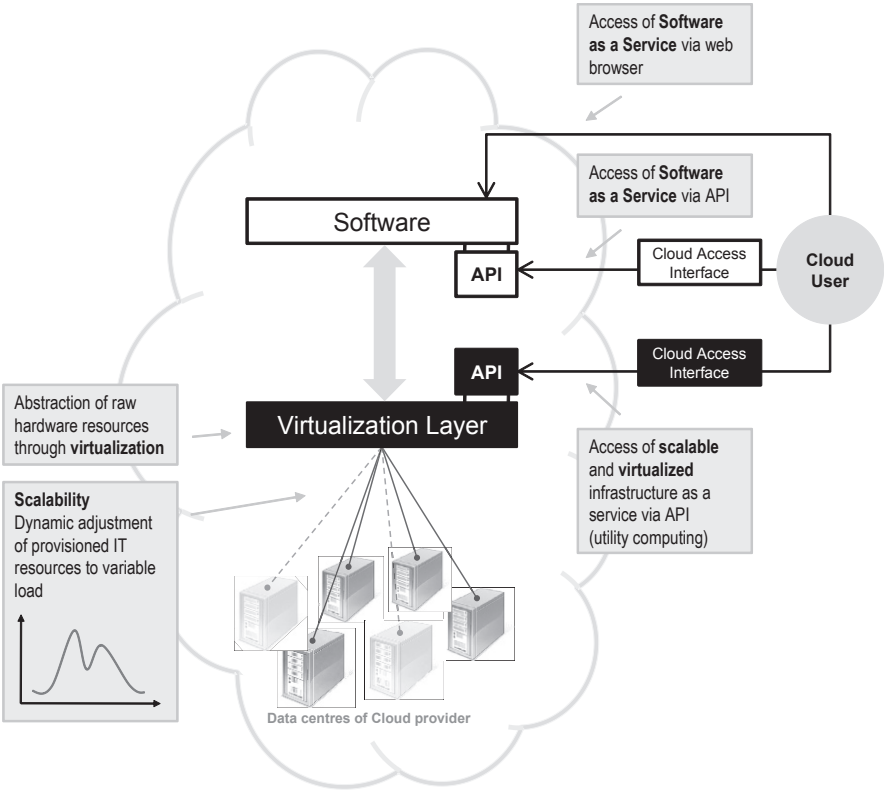

Fig. 4.1: Defining features of Cloud Computing

4.3 Architecture and Components of Clouds

In this section, we first provide an overview of concepts regarding the structure and components of Clouds. Then, we describe the most cited three-layer architectural concept for Clouds in detail.

4.3.1 Overview of Existing Concepts for Cloud Structures and Components

It is possible to find a number of concepts for Cloud structures in literature. At first sight, these classifications appear to differ from each other to varying extent. Eventually, however, they classify and describe the same phenomenon and share a common denominator.

Menken (2008) provides a very detailed concept consisting of 7 major components of Cloud Computing, namely application, client, infrastructure, platform, service, storage, and processing power. Miller (2008) looks at "different ways a company can use cloud computing to develop its own business applications", and distinguishes four types of Cloud service development, namely Software as a Service, Platform as a Service, Web Services, and On-Demand Computing. On-Demand Computing, as Miller (2008) notes, is also referred to as utility computing. Youseff et al. (2008) distinguish five layers of Cloud Computing: Cloud application, Cloud software environment, Cloud software infrastructure, software kernel, and firmware/hardware.

Forrester Research relate the components of Clouds to markets and distinguish five Cloud services markets. Two of them, Web-based services and SaaS offerings, are reported to be known markets that are delivered from the Cloud, whereas three cloud-infrastructure-as-a-service markets are new: app-components-as-a-service, software-platform-as-a-service, and virtual-infrastructure-as-a-service (Gilles et al. 2008). Finally, Reese (2009) considers SaaS as the term for "software in the cloud" and distinguishes four Cloud Infrastructure Models, namely Platform as a Service, Infrastructure as a Service, Private Clouds, and a fourth model representing all aspects of the previous Cloud infrastructure models.

All of the concepts above are very detailed and are influenced by the specific perspective on Clouds the respective authors take. Some of the concepts also involve aspects as Private Clouds and have different levels of detail for components that make up one logical entity. Given this, the concepts above do not provide a sufficiently generic description of a Cloud structure and its components. The concept most commonly used to describe a generic structure and components of Clouds is a 3-layered concept, which will be described in more detail in the next section.

4.3.2 The Three Layers of Cloud Computing

The definitions provided in section 4.2 already show that Cloud Computing comprises different IT capabilities, namely *infrastructure*, *platforms* and *software*. This may also be referred to as different 'shapes', 'segments', 'styles', 'types', 'levels' or 'layers' of Cloud Computing. Instead of speaking of different 'capabilities', thinking of it as different 'layers' makes much more sense because *infrastructure*,

platforms and *software* build subsequently upon the forerunning level and are logically connected as different layers of a Cloud architecture. Regardless of which term used, this threefold classification of Cloud Computing has become commonplace (Eymann 2008, Merrill Lynch 2008, O'Reilly 2008, RightScale 2008, Sun 2009a, Vaquero et al. 2008).

As the delivery of IT resources or capabilities as a service is an important characteristic of Cloud Computing, the three architectural layers of Cloud Computing are (see also fig. 4.2):

1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service (SaaS)

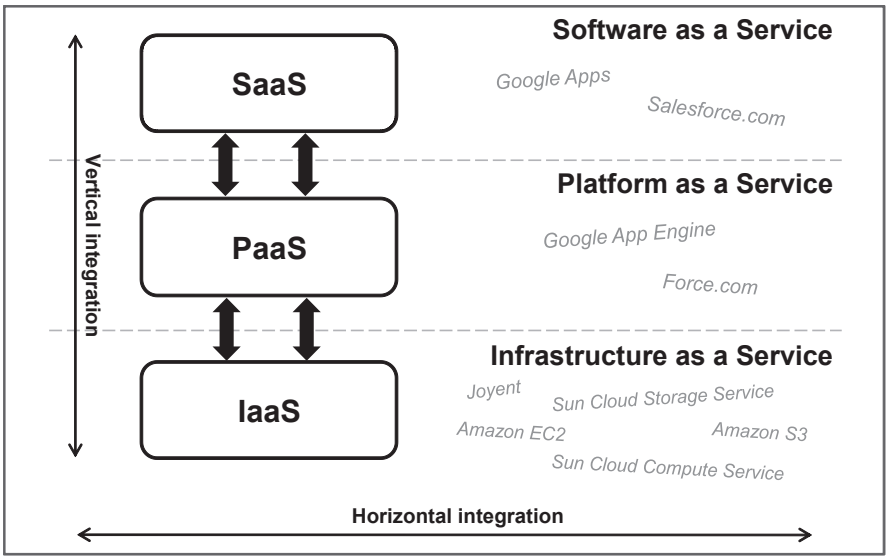

Fig. 4.2: The 3 layers of Cloud Computing: SaaS, PaaS, and IaaS

In the following subsections, we describe the three layers of Cloud Computing IaaS, PaaS and SaaS and how they are logically connected to each other.

4.3.2.1 Infrastructure as a Service (IaaS)

IaaS offerings are computing resources such as processing or storage which can be obtained as a service. Examples are Amazon Web Services with its Elastic Compute Cloud (EC2) for processing and Simple Storage Service (S3) for storage and Joyent who provide a highly scalable on-demand infrastructure for running Web sites and rich Web applications (Sun 2009a). PaaS and SaaS providers can draw upon IaaS offerings based on standardized interfaces. Instead of selling raw hardware infrastructure, IaaS providers typically offer virtualised infrastructure as a service. Foster et al. (2008) denote the level of raw hardware resources, such as compute, storage

and network resources, as the fabric layer. Typically by virtualization, hardware level resources are abstracted and encapsulated and can thus be exposed to upper layer and end users through a standardized interface as unified resources (Foster et al. 2008) in the form of IaaS (see figure 4.3).

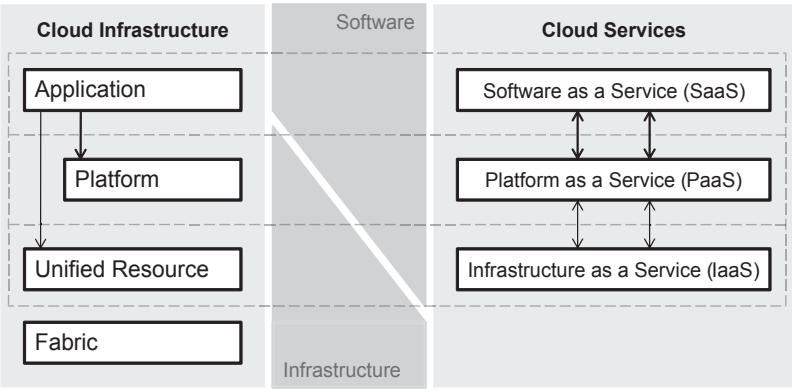

Fig. 4.3: Cloud Architecture related to Cloud services (adapted from Foster et al. 2008)

Already before the advent of Cloud Computing, infrastructure had been available as a service for quite some time. This has been referred to as utility computing, which is also used by some authors to denote the infrastructure layer of Cloud Computing (e.g. Armbrust et al. 2009, Miller 2008, O'Reilly 2008). Sun, for example, launched its Sun Grid Compute Utility in March 2006 (Schwartz 2006). The Sun Grid Compute Utility allowed users to purchase computing capability for $1/cpu-hr, i.e. on a pay-per-use basis. The Sun Grid Compute Utility could be accessed via Network.com. One year later, in March 2007, Sun announced the Network.com Application Catalog, which allowed developers and open source communities to just "click and run" their applications online (Sun 2007). Two years later, in March 2009, Sun announced its Open Cloud Platform as well as plans for its Sun Cloud, whose main services will be the Sun Cloud Storage Service and Sun Cloud Compute Service (Sun 2009b). Network.com, which once was the access point to the Sun Grid Compute Utility and the Network.com Application Catalog, was in a transition mode in early 2009 and now redirects to 'Sun Cloud Computing'(Sun 2009c, Sun 2009d).

Compared to the early utility computing offerings, IaaS denotes its evolution towards integrated support for all three layers (IaaS, PaaS, and SaaS) within a Cloud (see also Fellows 2009). From the early offerings of utility computing it became clear that for utility computing providers to be successful, they need to provide an interface that is easy to access, understand, program, and use, i.e. an API that would enable easy integration with the infrastructure of potential customers and potential developers of SaaS applications. Utility Computing providers' data centres are sufficiently utilized only if they are used by a critical mass of customers and SaaS providers.

As a consequence of the requirement for an easy and abstracted access to the physical layer of a Cloud, virtualization of the physical layer and programming platforms for developers emerged as major features of Clouds.

4.3.2.2 Platform as a Service (PaaS)

Platforms are an abstraction layer between the software applications (SaaS) and the virtualized infrastructure (IaaS). PaaS offerings are targeted at software developers. Developers can write their applications according to the specifications of a particular platform without needing to worry about the underlying hardware infrastructure (IaaS). Developers upload their application code to a platform, which then typically manages the automatic upscaling when the usage of the application grows (RightScale 2008). PaaS offerings can cover all phases of software development or may be specialized around a specific area like content management (Sun 2009a). Examples are the Google App Engine, which allows applications to be run on Google's infrastructure, and Salesforce's Force.com platform. The PaaS layer of a Cloud relies on the standardized interface of the IaaS layer that virtualizes the access to the available resources and it provides standardized interfaces and a development platform for the SaaS layer.

4.3.2.3 Software as a Service (SaaS)

As explained in section 3.6.2 in chapter 3, SaaS is software that is owned, delivered and managed remotely by one or more providers and that is offered in a pay-per-use manner (see also Mertz 2007). SaaS is the most visible layer of Cloud Computing for end-users, because it is about the actual software applications that are accessed and used.

From the perspective of the user, obtaining software as a service is mainly motivated by cost advantages due to the utility-based payment model, i.e. no up-front infrastructure investment. Well known examples for SaaS offerings are Salesforce.com and Google Apps such as Google Mail and Google Docs and Spreadsheets.

The typical user of a SaaS offering usually has neither knowledge nor control about the underlying infrastructure (Eymann 2008), be it the software platform which the SaaS offering is based on (PaaS) or the actual hardware infrastructure (IaaS). However, these layers are very relevant for the SaaS provider because they are necessary and can be outsourced. For example, a SaaS application can be developed on an existing platform and run on infrastructure of a third party. Obtaining platforms as well as infrastructure as a service is attractive for SaaS providers as it can alleviate them from heavy license or infrastructure investment costs and keeps them flexible. It also allows them to focus on their core competencies. This is similar to the benefits that motivate SaaS users to obtain software as a service.

According to market analysts, the growing openness of companies for SaaS and the high pressure to reduce IT costs are major drivers for a high demand and growth of SaaS, and by that also for Cloud Computing, in the next years. In August 2007, analyst firm Gartner forecasted an average annual growth rate of worldwide SaaS revenue for enterprise application software of 22.1% through 2011, reaching a

volume of $11.5 billion (Mertz et al. 2007). Analyst firm IDC estimates the growth rate of SaaS revenue to be 31% in 2009, which is more than four times of the total software market's growth rate (IDC 2008c). In October 2008, Gartner updated the estimates stating world wide SaaS revenue for enterprise application software is expected to more than double by 2012, reaching $14.5 billion (Gartner 2008c).

4.4 Opportunities and Challenges of Cloud Computing

As described in previous sections, Cloud Computing concerns the delivery of IT capabilities as a service on three levels: infrastructure (IaaS), platforms (PaaS), and software (SaaS). By providing interfaces on all three levels, Clouds address different types of customers:

- End consumers, who mainly use the services of the SaaS layer over a Web browser and basic offerings of the IaaS layer as for example storage for data resulting from the usage of the SaaS layer.
- Business customers that might access all three layers: the IaaS layer in order to enhance the own infrastructure with additional resources on demand, the PaaS layer in order to be able to run own applications in a Cloud and eventually the SaaS layer in order to take advantage of available applications offered as a service.
- Developers and Independent Software Vendors (ISVs) that develop applications that are supposed to be offered over the SaaS layer of a Cloud. Typically, they directly access the PaaS layer, and through the PaaS layer indirectly access the IaaS layer, and are present on the SaaS layer with their application.

In general, for all different kinds of Cloud customers, a Cloud offers the major opportunities known for X-as-a-Service offerings. From the perspective of the user, the utility-based payment model is considered as one of the main benefits of Cloud Computing. There is no need for up-front infrastructure investment: investment in software licenses and no risk of unused but paid software licenses, and investment in hardware infrastructure and related maintenance and staff. Thus, capital expenditure is turned into operational expenditure. Users of a Cloud service only use the volume of IT resources they actually need, and only pay for the volume of IT resources they actually use. At the same time, they take advantage of the scalability and flexibility of a Cloud. Cloud Computing enables easy and fast scaling of required computing resources on demand.

However, Cloud Computing has also several disadvantages: Clouds serve many different customers. Thus, users of a Cloud service do not know who else's job is running on the same server as their own ones (Sun 2009a). A typical Cloud is outside a company's or other organization's firewall. While this may not play a major role for consumers, it can have significant impact on a company's decision to move use Cloud Services. The major risks of Cloud Computing are summarized in table 4.2.

Table 4.2: Obstacles to adoption and growth of Cloud Computing

Obstacle	Source
Availability	Armbrust et al. (2009), IDC (2008a)
Security	IDC (2008a)
Performance	Armbrust et al. (2009), IDC (2008a)
Data lock-in	Armbrust et al. (2009)
Data confidentiality and auditability	Armbrust et al. (2009)
Data transfer bottlenecks	Armbrust et al. (2009)
Hard to integrate with in-house IT	IDC (2008a)
Lack of customizability	IDC (2008a)

The user has to rely on the promise of the Cloud provider with respect to reliability, performance and Quality of the Service (QoS) of the infrastructure. The usage of Clouds is associated also with higher security and privacy risks related to data storage and management in two ways: first because of the need to transfer data back and forth to a Cloud so that it can be processed in a Cloud; second because data is stored on an external infrastructure and the data owner relies on the Cloud provider's assurance that no unauthorized access takes place. Furthermore, the usage of Clouds requires an upfront investment in the integration of the own infrastructure and applications with a Cloud. At present, there are no standards for the IaaS, PaaS, and SaaS interfaces. This makes the choice of a Cloud provider and the investment in integration with Clouds risky. This can result in a strong log-in effect that is advantageous for the Cloud provider but disadvantageous for the users.

Given the risks associated with the usage of Clouds, in each case a careful evaluation and comparison of the potential benefits and risks is necessary. Also, it needs to be considered which data and processes are suitable to be used for "Cloud sourcing" and which should better be not exposed to any organization outside the firewall.

4.5 Classification of Clouds

Clouds can generally be classified according to who the owner of the Cloud data centres is. A Cloud environment can comprise either a single Cloud or multiple Clouds. Thus, it can be distinguished between single-Cloud environments and multiple-Cloud environments. The following subsections provide a classification of single-Cloud environments according to the Cloud data centre ownership (sec. 4.5.1) and a classification of multiple-Cloud environments according to which type of Clouds are combined (sec. 4.5.2).

4.5.1 Public Clouds vs. Private Clouds

In section 4.2, based on the review of many Cloud definitions, we have characterized Cloud Computing as the delivery of IT capabilities to external customers, or, from the perspective of a user, obtaining IT capabilities from an external provider,

as a service in a pay-per-use manner and over the Internet. In addition, we have identified scalability and virtualization as key characteristics of Cloud Computing. External data centres, e.g. those of Google or Amazon, are thus the foundation on the raw hardware or fabric level for delivering IT capabilities as Cloud services.

However, virtualizing raw hardware resources and offering them as abstracted IT capabilities as a service is not necessarily bound to the external delivery mode usually associated with Cloud Computing. Companies and other organizations also use virtualization and service-oriented computing to increase utilization of their existing IT resources and to increase flexibility. The utilization rate of traditional server environments is between 5 to 15% (e.g. IBM 2008). Increasing it to up to 18% is reported to be easily achievable (Lohr 2009, McKinsey 2009). Through aggressive virtualization, large companies can increase their server utilization rates to up to 35%, which is close to the level of Cloud providers such as Google with 38% (Lohr 2009, McKinsey 2009). Higher utilization makes possible to consolidate server environments, i.e. the number of physical servers can be reduced. This lowers hardware maintenance costs, required physical space for the servers, power and cooling costs as well as the carbon footprint of IT.

To distinguish between external providers of Cloud services (external Clouds) and companies' efforts to build internal Cloud infrastructures (internal Clouds) two distinct terms are commonly used: *Public* Cloud for external Clouds and *Private* Cloud for internal Clouds (see e.g. Armbrust et al. 2009, IBM 2009, Reese 2009, Sun 2009a).

A *Public Cloud* is data centre hardware and software run by third parties, e.g. Google and Amazon, which expose their services to companies and consumers via the Internet (Armbrust et al. 2009, IBM 2009, Sun 2009a). A Public Cloud is not restricted to a limited user base: it "…is made available in a pay-as-you-go manner to the general public" (Armbrust et al. 2009). Thus, Clouds can address two type of customers: either end consumers on the B2C market or companies on the B2B market.

Companies may not be willing to bear the risks associated with a move towards a Public Cloud and may therefore build internal Clouds in order to benefit from Cloud Computing. *Private Clouds* refer to such internal data centres of a company or other organization (Armbrust et al. 2009). A Private Cloud is fully owned by a single company who has total control over the applications run on the infrastructure, the place where they run, and the people or organizations using it – simply over every aspect of the infrastructure (Sun 2009a, Reese 2009). A Private Cloud relies on virtualization of an organization's existing infrastructure (Reese 2009), leading to benefits such as increased utilization as described above. The key advantage of a Private Cloud is to gain all advantages of virtualization, while retaining full control over the infrastructure (Reese 2009).

The definitions of Cloud Computing reviewed in section 4.2 clearly show that Cloud Computing concerns the delivery of IT capabilities to *external* customers, or, from the perspective of the user, obtaining IT capabilities from *external* providers. Thus, some authors do not consider Private Clouds, or *internal* Clouds, as part of or as true Cloud Computing (e.g. Armbrust et al. 2009, Reese 2009). Reese (2009), for

example, notes that Private Clouds lack "the freedom from capital investment and the virtually unlimited flexibility of cloud computing."

4.5.2 Hybrid Clouds and Federations of Clouds

Single Clouds can be combined resulting in multiple-Cloud environments. Contingent on which types of Clouds (public or private) are combined, two types of multiple-Cloud environments can be distinguished:

- Hybrid Clouds and
- Federation of Clouds.

Hybrid Clouds combine Public and Private Clouds and allow an organization to both run some applications on an internal Cloud infrastructure and others in a Public Cloud (Sun 2009a). This way, companies can benefit from scalable IT resources offered by external Cloud providers while keeping specific applications or data inside the firewall. A mixed Cloud environment adds complexity regarding the distribution of applications across different environments, monitoring of the internal and external infrastructure involved, security and privacy, and may therefore not be suited for applications requiring complex databases or synchronization (Sun 2009a).

The terms *Federated Clouds* or *Federation of Clouds* denote collaboration among mainly Public Clouds even though Private Clouds may be involved. Cloud infrastructure providers are supposed to provide massively scalable computing resources. This allows users and Cloud SaaS providers not to worry about the computational infrastructure required to run their services. The Cloud infrastructure providers, however, may face a scalability problem themselves. A single hosting company may not be able to provide seemingly infinite computing infrastructure, which is required to serve increasing numbers of applications, each with massive amounts of users and access at anytime from anywhere. Consequently, Cloud infrastructure providers may eventually partner to be able to truly serve the needs of Cloud service providers, i.e. providing seemingly infinite compute utility. Thus, *the* Cloud might become a federation of infrastructure providers or alternatively there might be a federation of clouds (RESERVOIR 2008).

Federated Clouds are a collection of single Clouds that can interoperate, i.e. exchange data and computing resources through defined interfaces. According to basic federation principles, in a Federation of Clouds each single Cloud remains independent, but can interoperate with other Clouds in the federation through standardized interfaces. At present, a Federation of Clouds seems still to be a theoretical concept as there is no common Cloud interoperability standard. One new initiative that tries to develop a common standard is the Open Cloud Computing Interface, which is developed by the Open Cloud Computing Interface Working Group (http://www.occi-wg.org/) of the Open Grid Forum (OGF). The goal is through a standardized API among Clouds to enable both interoperability among Clouds from different vendors and new business models and platforms as (according to OCCI 2009):

- Integrators, that offer advanced management services that spread over several Clouds or Hybrid Clouds
- Aggregators that offer a single common interface to multiple Cloud providers.

The integration and advances in interoperability of Clouds might be an important factor for the future success of Cloud Computing. Open standards and interoperability among Private and Public Clouds enable a higher flexibility for user companies. The user companies would be able to also partly outsource data and processes to the Cloud that are less security- and privacy-sensitive. At the same time, the possibility to build a Federation of Clouds would enable specialization of single Clouds as well as a broader choice for the users.

4.6 Grid and Cloud Computing Compared

The description of Grid Computing in Chapter 3 and Cloud Computing in this chapter show that there are many similarities among Grid and Cloud Computing. This has provoked many discussions in commercial and scientific literature around the question if Grids and Clouds are the same, if Cloud Computing is only a new marketing hype, or if there are substantial differences between Grid and Cloud Computing.

Currently, the discussion about differences among Grid and Cloud Computing mainly regards technical aspects (see also table 4.3).

Table 4.3: Grid and Cloud Computing technically compared

	Grid Computing	Cloud Computing
Means of utilisation (e.g. Harris 2008)	Allocation of multiple servers onto a single task or job	Virtualization of servers; one server to compute several tasks concurrently
Typical usage pattern (e.g. EGEE 2008)	Typically used for job execution, i.e. the execution of a programme for a limited time	More frequently used to support long-running services
Level of abstraction (e.g. Jha et al. 2008)	Expose high level of detail	Provide higher-level abstractions

Foster et al. (2008) for example identify differences among Grid and Cloud Computing in various aspects as security, programming model, compute model, data model, application and abstraction. According to Merrill Lynch (2008), what makes Cloud Computing new and differentiates it from Grid Computing is virtualization: “Cloud computing, unlike grid computing, leverages virtualization to maximize computing power. Virtualization, by separating the logical from the physical, resolves some of the challenges faced by grid computing” (Merrill Lynch 2008). While Grid Computing achieves high utilization through the allocation of multiple servers onto a single task or job, the virtualization of servers in Cloud Computing achieves high utilization by allowing one server to compute several tasks concurrently (Harris 2008). Beside these technological differences between Grid and

Cloud, there are differences in the typical usage pattern. Grid is typically used for job execution, e.g. the execution of a HPC programme for a limited time. Clouds do support a job usage pattern but are more frequently used to support long-running services (EGEE 2008).

While most authors acknowledge similarities among those two paradigms, the opinions seem to cluster around the statement that Cloud Computing has evolved from Grid Computing and that Grid Computing is the foundation for Cloud Computing. Foster et al. (2008) for example describe the relationship between Grid and Cloud Computing as follows:

> "We argue that Cloud Computing not only overlaps with Grid Computing, it is indeed evolved out of Grid Computing and relies on Grid Computing as its backbone and infrastructure support. The evolution has been a result of a shift in focus from an infrastructure that delivers storage and compute resources (such is the case in Grids) to one that is economy based aiming to deliver more abstract resources and services (such is the case in Clouds)."

Thus, Cloud and Grid computing can be considered as complementary. Grid interfaces and protocols can enable the interoperability between resources of Cloud infrastructure providers and/or a Federation of Clouds. Grid solutions for job computing can run as a service on top of a Federation of Clouds and/or a distributed virtualized infrastructure (Llorente 2008a, Llorente 2008b). In addition, the potential benefits of simplicity offered by Cloud technologies, such as higher-level of abstractions (Jha et al. 2008), may help to better serve current Grid users, "attract new user communities, accelerate grid adoption and importantly reduce operations costs" (EGEE 2008).

In the discussion about the differences among Grids and Clouds, less attention is given to explaining them from user perspective yet. Based on the described features of Grid Computing in chapter 3 and Cloud Computing in this chapter, the main changes from the user perspective can be summarized as follows:

- *Pure focus on X-as-a-Service (XaaS) by Clouds*: As mentioned in section 3.2 in chapter 3, the basis for Grid Computing is Grid middleware that is available on the market as packaged or open source software. Utility Computing is only one form of Grid Computing. Compared to that, Cloud Computing focuses purely on XaaS offered in a pay-per-use manner. There is no middleware that enables the building of Clouds yet.
- *Focus on different types of applications:* Grid Computing emerged in eScience to solve scientific problems requiring HPC. Current usage in industry also focuses mainly on HPC, for example in collaborative engineering based on simulation, in research and development in pharmaceutical companies and similar. HPC applications are usually batch-oriented and require high computing power for one task that is run once in a time. Given this, Grid Computing has the goal to assign computing resources, in many cases from different domains, to such HPC tasks. Cloud Computing is rather oriented towards applications that run permanently (e.g. the well-known CRM SaaS Salesforce.com) and have varying demand for physical resources while running. In order to be more flexible, one

major difference of Cloud Computing to Grid Computing is virtualization and adjustment of provided resources to demand. Thus, Cloud Computing extends the spectrum to which virtualization can be applied.

- *Different relationships among resource providers:* The goal of Grid Computing is creation of VOs with clear up-front commitment of the involved parties and encoding of agreements and polices in the software. Cloud Computing eliminates the need for an up-front commitment by Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs (see also Armbrust et al. 2009).
- *Different scope of offerings:* Grid Computing clearly focuses on providing infrastructure as a service, or utility computing. Cloud Computing provides an integrated support for IaaS, PaaS and SaaS. Given this, Cloud Computing makes the development of SaaS applications easier.
- *Extended scope of interfaces to the user:* Grid Computing allocates heterogeneous resources to one task and focuses on communication among different resources on the physical layer and towards the application running on it. The Grid interfaces are rather based on protocols and APIs and by that only usable by technical experts. Cloud Computing is designed to provide interfaces for end users over Web browser or through APIs. Thereby there are different and specific APIs on each layer (IaaS, PaaS, and SaaS). Given the higher level of abstraction and the different interfaces, Cloud Computing is suitable to address end users in the B2C and C2B market at the same time.

To summarize, Grid Computing provides the means to share and unify heterogeneous computing resources. It is the starting point and basis for Cloud Computing. Cloud Computing essentially represents the increasing trend towards the external deployment of IT resources, such as computational power, storage or business applications, and obtaining them as services.