Cloud Computing

Eine Abgrenzung, Geschäftsmodelle und Forschungsgebiete

Die Autoren

Prof. Dr. Christof Weinhardt
Dipl.-Inform.Wirt Arun Anandasivam
Dipl.-Inform.Wirt Benjamin Blau
Dipl.-Inform. Nikolay Borissov
Dipl.-Math. Thomas Meinl
Dipl.-Inform.Wirt Wibke Michalk
Dr. Jochen Stößer

Autorenadresse zur Veröffentlichung

Organisation: Universität Karlsruhe (TH)
Organisationseinheit: Institut für Informationswirtschaft und –management (IISM)
Straße: Englerstr. 14
PLZ Ort (PLZ ohne Länderbuchstaben): 76131 Karlsruhe
evtl. Telefon im DIN-Format: 0721 608-8370
E-Mail: nachname@iism.uni-karlsruhe.de
Web: http://www.im.uni-karlsruhe.de

Cloud Computing

Eine Abgrenzung, Geschäftsmodelle und Forschungsgebiete

Zusammenfassung

In letzter Zeit hat ein neues Rechen- und Architekturparadigma Einzug gehalten, das sogenannte Cloud Computing. Es wird mindestens so umfangreich beworben wie Grid Computing wenige Jahre zuvor, was u.a. zahlreiche Diskussionen über die Abgrenzung zwischen Grid und Cloud Computing angeregt hat. Der erste Beitrag dieses Artikels ist deshalb die eingehende Diskussion der verschiedenen Charakteristika von jeweils Grid und Cloud Computing. Diese technische Abgrenzung ermöglicht eine fundierte Diskussion der Geschäftschancen des Cloud Computing Paradigmas. In diesem Artikel wird hierzu zunächst ein Framework für Geschäftsmodelle in Clouds aufgestellt. Anschließend werden aktuelle Cloud-Angebote besprochen und in das Framework eingereiht. Zum Schluss spricht dieser Artikel Herausforderungen und erfolgversprechende Forschungsgebiete an, die adressiert und bewältigt werden müssen, damit die Vision von Clouds erfolgreich umgesetzt werden kann.

Stichworte

Cloud Computing, Grid Computing, Geschäftsmodelle, Forschungsgebiete

Cloud Computing

A Classification, Business Models, and Research Directions

Abstract

Lately, a new computing paradigm has emerged: “Cloud Computing”. It seems to be promoted as heavily as the “Grid” was a few years ago, causing broad discussions on the differences between Grid and Cloud Computing. The first contribution of this paper is thus a detailed discussion about the different characteristics of Grid Computing and Cloud Computing. This technical classification allows for a well-founded discussion of the business opportunities of the Cloud Computing paradigm. To this end, this paper first presents a business model framework for Clouds. It subsequently reviews and classifies current Cloud offerings in the light of this framework. Finally, this paper discusses challenges that have to be mastered in order to make the Cloud vision come true and points to promising areas for future research.

Keywords

Cloud Computing, Grid Computing, Business Models, Research Directions

Vorspann als Fließtext ohne Spiegelstriche

Zunächst nimmt dieser Artikel eine technische Abgrenzung von Grid und Cloud Computing vor. Dies ermöglicht eine fundierte Diskussion der Geschäftschancen von Cloud Computing. Zunächst wird hierzu ein Framework für Geschäftsmodelle in Clouds vorgestellt. Mit Hilfe dieses Frameworks lassen sich Geschäftsmodelle auf der Ebene der Infrastruktur, der Plattform oder der Applikation ansiedeln. Anschließend werden aktuelle reale Cloud-Dienste besprochen, in das Framework eingeordnet sowie nach dem Zahlungsschema und dem Typ des Dienstes kategorisiert. Zum Schluss spricht dieser Artikel Herausforderungen und erfolgversprechende Forschungsgebiete an, die adressiert und bewältigt werden müssen, damit sich die Vision von Clouds erfüllen kann, wie z.B. die Entwicklung einer Cloud API, die Bepreisung von komplexen, zusammengesetzten, sprich komponierten Diensten, sowie Sicherheit und Vertrauenswürdigkeit.

1 Introduction

Recent years' headlines in newspapers and magazines promoted "Grid" as (one of) the most promising trends in the IT sector. Arising from the need for computational power that could not be provided by clusters, distributed high performance computing in virtual organizations enabled researchers to deal with large amounts of data. New challenges arose, e.g., for the CERN[1] when planning the experiments in the Large Hadron Collider (LHC)[2]. The CERN needed the ability of high-performance computing for dealing with petabytes of data being created by the LHC. Distributed computing projects like distributed.net[3] that connects users to a network for searching keys to encryption algorithms were a basis for Grid computing. Expanding these first approaches, the EGEE project was founded in order to provide researchers with a Grid infrastructure and hence with compute power independent of their location. Later on, the profitability of Grids for other institutions became obvious, as it was no longer necessary to buy all the computing resources and host a large server farm but it was possible to use the Grid whenever the resources where needed.

Discovering these facts, research projects were started all over the world, being funded by governments as well as by the industry, examining different aspects important to Grid Computing. An example is SORMA[4], developing a market-based mechanism for resource-allocation in a decentralized environment (Neumann et al. 2008).

Finally, IBM adopted the Grid technology and implemented "Grid and Grow", a Grid solution that is easy to deploy in order to enable even those who are not specialized in computer science to start using Grid technology. Oracle offers the possibility to dynamically add capacity to ones computing power by using Oracle Grid[5], offering a business case for the Grid.

But lately, a new computing paradigm has emerged: "Cloud Computing". It seems to be promoted as "Grid" has been a few years ago, giving way to broad discussions on the differences between Grid and Cloud Computing. As there is a lot of confusion on the definition of Cloud Computing, the following sections will try to provide guidance concerning the questions if Cloud Computing is simply a renaming of already known technologies or if it paves the way for a commercially wide-spread usage of large-scale IT resources.

This paper's contribution is threefold. First, a list of criteria for distinguishing between Grids and Clouds is presented. Afterwards, business models ensuring Clouds' sustainability are illustrated and mapped to current Cloud offerings. Finally, future work in the field of Cloud Computing is described pointing to new research directions.

2 Grid vs. Cloud Computing

The objective of this section is to distinguish the concepts of Grid Computing and Cloud Computing. While we do not aim at giving a clear and unambiguous definition of Cloud Computing (maybe this term was meant to be foggy after all), we do give a set of criteria that might help in at least roughly delineating the boundaries between these concepts.

2.1 Grid Computing

In the mid 1990s, the term "Grid Computing" was derived from the electrical power Grid to emphasize its characteristics like pervasiveness, simplicity and reliability (Foster, Kesselman, 1999). The upcoming demand for large-scale scientific applications required more computing power than a cluster within a single domain (e.g. an institute) could provide. Due to the fast interconnectedness via the Internet, scientific institutes were able to share and aggregate geographically distributed resources including cluster systems, data storage facilities and data sources owned by different organizations. However, resource sharing among distributed

[1] http://www.cern.de/

[2] http://lhc.web.cern.ch/lhc/

[3] http://www.distributed.net/

[4] http://www.sorma-project.org/

[5] http://www.oracle.com/technologies/Grid/

systems by applying standard protocols and standard software was rarely commercially realized (e.g. SUN “N1 Grid Engine” or IBM “Grid and Grow”). The development of Grid Computing and its standards was mainly driven by scientific Grid Communities. The major community is the Open Grid Forum[6]. The Open Grid Forum (OGF) is consolidating and leading the standardization initiatives within the Grid community. Ian Foster as an active member in the Grid Community came up with a “three point checklist” (Foster 2002), which has become the most widely used definition. For Foster (Foster 2002), Grid Computing is characterized by

- *decentralized resource control*, i.e. the Grid resources are locally dispersed and span multiple administrative domains;
- *standardization*, i.e. the Grid middleware is based upon open and common protocols and interfaces;
- *non-trivial qualities of service*, e.g. regarding latency, throughput, and reliability.

While decentralized control differentiates Grid Computing from cluster computing, Grid-like systems such as BOINC[7] (formerly known as Seti@home) or Folding@home[8] typically do not rely on a standardized middleware. Non-triviality is oftentimes extended to cover the nature of the application; Grid applications typically go beyond simple file sharing as in peer-to-peer systems. These Grid-like systems use specific software developed for the project and do not allow any flexibility for extension and calculations for different kinds of computational jobs. Furthermore, the results of the jobs are merged together at one institute (centralized system).

Foster et al. (2001) outline that resource sharing among different administrative domains has to be governed by certain organizational structures and rules. Hence, participating institutes are typically managed in virtual organizations, where institutes agree on common sharing rules.

Grid applications require a middleware to enable communication over a standardized protocol with the computing and data resources. The most prominent Grid middlewares currently are Globus Toolkit 4[9], Unicore[10] and gLite[11]. All Grid middlewares attempt to apply the standards defined by the OGF community. One of the important standards is the broad specification of the architecture called Open Grid Services Architecture (OGSA). OGSA is not a single specification, but rather a set of related standards including Basic Execution Services, Job Submission Description Language, and Data Access and Integration Services (IBM 2006).

2.2 Cloud Computing

For Cloud Computing, there is no established definition yet, which contributes to the current skepticism and / or possible overestimation regarding its impact on the technology and business landscape.

For Boss et al. (2007), “a Cloud is a pool of *virtualized* computer resources”. They consider Clouds to complement Grid environments by supporting the *management* of Grid resources. In particular, according to Boss et al. (2007), Clouds allow

- the dynamic scale-in and scale-out of applications by the provisioning and de-provisioning of resources, e.g. by means of virtualization;

[6] http://www.ogf.org/
[7] http://boinc.berkeley.edu/ and http://setiathome.berkeley.edu/
[8] http://folding.stanford.edu/
[9] http://www.globus.org/
[10] http://www.unicore.eu/
[11] http://glite.web.cern.ch/

- the monitoring of resource utilization to support dynamic load-balancing and re-allocations of applications and resources.

Most importantly, Boss et al. (2007) stress that Clouds are not limited to Grid environments, but also support "*interactive, user-facing applications*" such as Web applications and three-tier architectures.

Lawton (2008) briefly describes the type of applications that is run in Clouds: *Web-based applications* that are accessed via browsers but with the look-and-feel of desktop programs. While this focus might be a bit too narrow, this corresponds to the type of applications that currently emerge on top of Amazon's Elastic Compute Cloud (EC2), which are oftentimes Web 2.0 applications that need to grow and scale quickly[12].

For Skillicorn (2002), Cloud Computing implies "*component-based application construction*". Instead of having to develop applications entirely from scratch, application fragments such as simple (Web) services, third-party software libraries etc. can be dynamically retrieved from and assembled in the Cloud. This corresponds to the additional services that are offered by Amazon, such as Simple Storage and SimpleDB[13], or the former model of Sun Microsystem's network.com platform[14], which offered so-called "published applications" for re-use.

For Weiss (2007), Cloud Computing is not a fundamentally new paradigm. It draws on existing technologies and approaches, such as Utility Computing, Software-as-a-Service, distributed computing, and centralized data centers. What is new is that Cloud Computing combines and integrates these approaches. Especially the combination with *Utility Computing* and *data centers* seems to differentiate Cloud Computing from Grid Computing. While Utility Computing has already been proposed earlier (e.g. Rappa 2004) and is in principle also applicable to Grid Computing, up to now business models and pricing only became accepted and implemented in the context of Cloud Computing. Though there do exist research projects (like the SORMA[15], Biz2Grid[16] and GridEcon[17] projects) developing mechanisms and software to realize these concepts in Grid environments as well, those topics are mostly frowned upon especially in science Grids. Moreover, while Grid Computing is partly defined by its dispersed resources and decentralized control, Cloud Computing seems to be a step back towards centralizing IT in data centers again to economize on scale and scope.

Speaking of Software-as-a-Service, one naturally arrives at the term *Application Service Provider* (ASP), which became popular about a decade ago. Not receiving much attention any more these days, the similarities to the concept of Cloud Computing are evident: The ASPs already implemented business models and pricing, allowing a similar acquisition of software as Clouds do today. However, the services offered in a Cloud include more like simple hardware and complex (mashed-up) services, thus already excelling the former ASP concept. Furthermore, while ASPs never really got in the business at large scale, matters are different with Cloud Computing, which – despite its relatively short appearance – has already established a stable number of customers, and this number is growing rapidly. This is in parts due to improved technical frameworks, including transmission techniques and security issues.

Similar to Weiss et al. (2007), Gentzsch (2008) states that Grids have not had sustainable business models thus far but were largely based on public funding. To Gentzsch, similar to Boss et al. (2007), Clouds are a "useful utility that you can plug into your Grid".

[12] http://aws.amazon.com/solutions/case-studies/
[13] http://www.amazon.com/simpledb
[14] http://www.network.com/
[15] http://sorma-project.org/
[16] http://www.biz2grid.de/
[17] http://www.gridecon.eu/

2.3 A Criteria Catalogue

Table 1 summarizes the key differences between Grid and Cloud Computing, which can be derived from theoretical concepts in the literature as well as already established Grid or Cloud implementations.

Tab. 1 Grid vs. Cloud Computing

Criteria	Grid Computing	Cloud Computing	Literature / Practical Example
Virtualization	in its beginning	essential	Boss et al. (2007), Amazon EC2
Type of application	batch	interactive	Boss et al. (2007), Lawton (2008), Amazon EC2
Development of applications	local	in the Cloud	Skillicorn (2002), Sun network.com
Access	via Grid middleware	via standard Web protocols	Lawton (2008), Amazon EC2
Organizations	virtual	physical	
Business Models	sharing	pricing (utility model)	Weiss (2007), Gentzsch (2008)
SLAs/Liability	not yet enforceable	essential	
Control	decentralized	centralized (data center)	Weiss (2007), Foster (2002)
Openness	high	low	
Ease of use	hard (up to now)	easy	
Switching Cost	low due to standardization	high due to incompatibilities	

While the use of virtualization technologies (e.g. Xen[18], VMWare[19]) in Grid environments is just in its infancy, the abstraction from the underlying hardware pool is essential for Cloud solutions. Virtualization technology allows Cloud providers to run multiple so-called virtual machines on one physical machine. This is transparent to the application and its end-user and further has the benefit that each machine can be custom-tailored towards the needs of the users, e.g. with respect to the technical requirements and pre-installed software libraries. From the provider's point of view, virtualization enables the efficient utilization (e.g. with respect to load-balancing and energy-consumption) of physical machines. Moreover, a virtual machine essentially constitutes a sandbox, which prevents side-effects between applications and users and prevents applications from compromising the physical resources.

The vast majority of Grids is employed to harness idle computing capacity for large computational jobs. As pointed out above, Clouds are already used for very different kinds of applications, such as the hosting of Web sites. In Grids, only the so-called head nodes of clusters are visible to the outside and the end-user does not have direct access to the resources. In Amazon EC2, each virtual machine is assigned an IP address. The user can thus directly connect to her machine enabling the hosting of all kinds of interactive applications.

[18] http://www.xen.org/
[19] http://www.vmware.com/

Furthermore, the approach to the development of applications is very different in Grids and Clouds. In Grids, the user typically needs to generate an executable file of her application. This executable is then transferred to and executed on the remote resources in the Grid. Clouds allow a fundamentally different approach to software development. For instance, the network.com platform of Sun Microsystems and force.com by Salesforce.com offer ready-to-use application components. The vision is that the user can dynamically assemble these existing functionalities to a full application on the platform as such.

At the moment, access to Grid resources is realized via a specific, sometimes very complex middleware, required to run on both the client's as well as on the provider's side. In contrast, as discussed above, interaction with resources in the Cloud are established via standard Web protocols, facilitating the access for the users. The initial participation in Grid Computing networks and thus the upfront investment can be very high, which may be one of the reasons why Grid Computing has not yet been established successfully in many business scenarios. The lightweight accessibility and ease of use is one key factor why Cloud vendors currently succeed in boosting their amount of non-academic customers in a relative short period of time.

While virtual organizations have become a common instrument in Grid environments to enable trust between all participants, organizational structures in Clouds remain on the physical level, as to date all Cloud Computing offerings can be mapped to individual companies, without hardly any mergers. This results in Grids being quite open and, from an organizational point of view, easily accessible for new participants, while it is nearly impossible for any interested party to join a Cloud infrastructure of an already established Cloud vendor.

This principal point of different organizational realizations also results in very different switching costs: While for any Grid user switching from the resources of one Grid provider to another is relatively easy, due to the lack of standards in Clouds, this does not hold for Cloud Computing settings. Typically, Cloud providers have no interest in participating and implementing standards enabling potential customers to switch easily and, thus, are maintaining proprietary interfaces and infrastructures. Being a downside for the clients in the first place, this however enables Cloud providers to offer and sign Service Level Agreements (SLAs) with their customers, encouraging the use of Clouds even for mission-critical industrial services, as these SLAs with one single provider are enforceable. Today SLA enforcement in Grid environment is still in research stages, as with the submission of a Grid resource request the allocated provider is generally not known a priori, this requires a dynamic automated SLA contracting and enforcement scheme, which proves hard to be realized in the Grids field due to its many complex aspects.

3 Cloud Business Models

Grid Computing has its roots in the eScience communities and never really gained much commercial attention. Grids are mainly used in the field of physics, biology and other computing intense applications in the research sector. The fact that Grids are facilitated by small groups of specialized experts implies that there has never been the need for a refinement and adaption of the Grid concept to make it utilizable and attractive for business-oriented and commercial communities. Grids in practice lack rudimentary components that are inevitable for the applicability in real-world business scenarios such as comprehensive service-level-management concepts that provide service-level negotiation, monitoring and enforcement, components for market-based allocation of Grid resources, dynamic pricing and monetary compensation mechanisms. Current trends in Cloud Computing expose a strong ambition to close these gaps and to establish existing concepts and technologies within the business world. Consequently, these trends motivate companies to incorporate innovative business models focusing on various aspects of Cloud Computing. In order to attain a better understanding and a common conceptualization, we developed a Cloud Business Model Framework (Fig. ***1***) that provides a hierarchical classification of different business models and some well-known representatives within the Cloud.

3.1 A Cloud Business Model Framework

The Cloud Business Model Framework (CBMF) is mainly categorized in three layers, analogously to the technical layers in Cloud realizations, such as the *infrastructure layer*, the *platform-as-a-service layer* and the *application layer* on top.

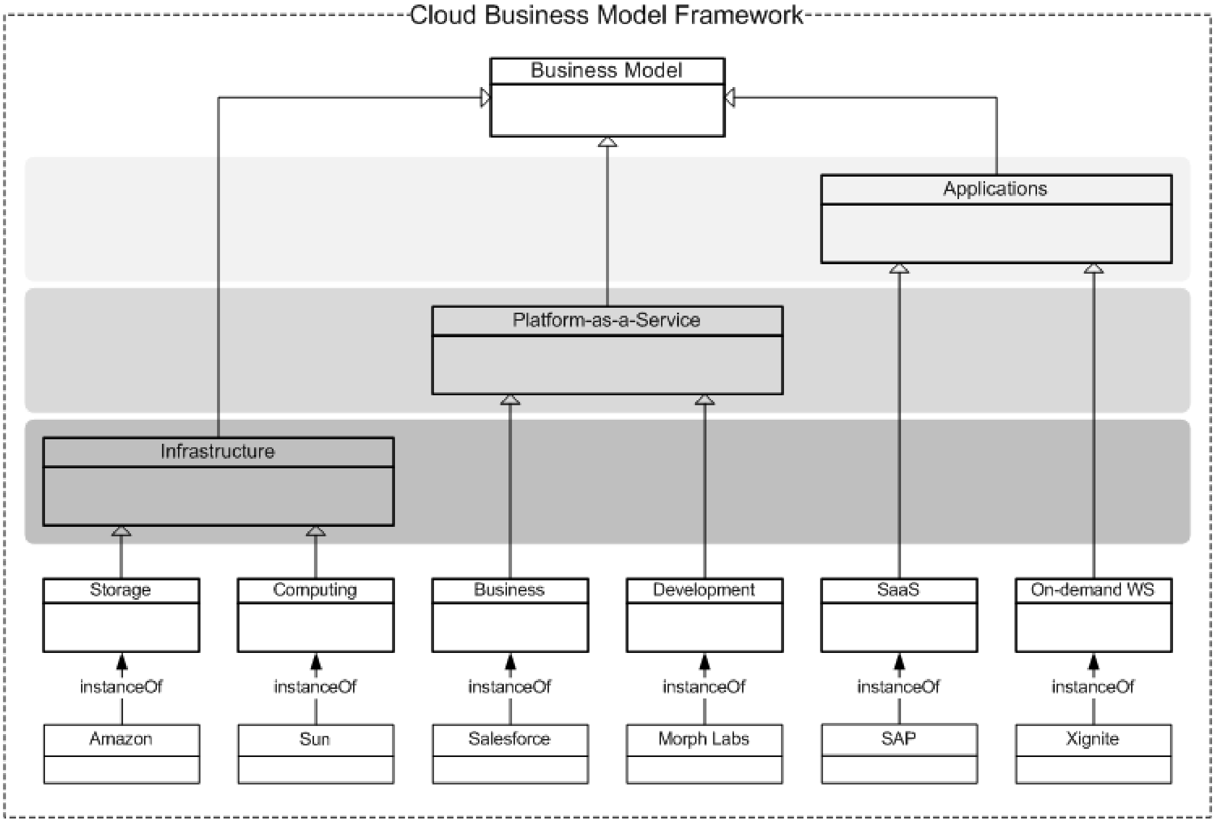

Fig. 1 Cloud Business Model Framework

Infrastructures in the Cloud – The infrastructure layer comprehends business models that focus on providing enabler technologies as basic components for cloud computing ecosystems[20]. We distinguish between two categories of infrastructure business models: The provision of storage capabilities and the provisioning of computing power. For example, Amazon offers services based on their infrastructure as a computing service (EC2[21]) and a storage service (S3[22]). So far, pricing models are mostly pay-per-use or subscription-based. In most cases, Cloud Computing infrastructures are organized in a cluster-like structure facilitating virtualization technologies. Among providing pure resource services, providers such as RightScale[23] often enrich their offerings through value-added services for managing the underlying hardware (i.e. scaling, migration) that are accessible via Web frontends.

Platforms in the Cloud – This layer represents platform solutions on-top of a cloud infrastructure that provides value-added services (platform-as-a-service) from a technical and a business perspective. We distinguish between development platforms and business platforms. Development platforms enable developers to write their applications and upload their code into the cloud where the application is accessible and can be run in a web-based manner. Developers do not have to care about issues like system scalability as the usage of their ap-

[20] We refer to the term cloud computing ecosystem as the fruitful interplay and coopetition between all players that realize different business models in the cloud computing context.

[21] http://aws.amazon.com/ec2/

[22] http://aws.amazon.com/s3/

[23] http://www.rightscale.com/

plications grows. Prominent examples are Morph Labs[24] and Google App Engine[25] which provide platforms for the deployment and management of Grails, Ruby on Rails and Java applications in the cloud. A further example is BungeeLabs[26] that provide a platform that provides functionality for managing the whole web application lifecycle from development to productive provisioning. Business platforms such as Salesforce[27] with their programming language Apex and Microsoft with their business platform xRM (that is still in a development phase) have also gained strong attention and enable the development, deployment and management of tailored business applications in the cloud.

Applications in the Cloud – the application layer is what most people get to know from Cloud Computing as it represents the actual interface for the customer. Applications are delivered through the Cloud facilitating the platform and infrastructure layer below which are opaque for the user. We distinguish between Software-as-a-Service (SaaS) applications and the provisioning of rudimentary Web services on-demand. Most prominent examples in the SaaS area are Google Apps[28] with their broad catalogue of office applications such as word and spreadsheet processing as well as mail and calendar applications that are entirely accessible through a web browser. An example from the B2B sector is SAP that delivers their service-oriented business solution BusinessByDesign[29] over the web for a monthly fee per user. In the field of Web service on-demand provisioning well-established examples are Xignite[30] and StrikeIron[31] that offer Web services hosted on a Cloud on a pay-per-use basis.

3.2 Cloud Computing Offerings

Currently, we are observing a rising number of offerings of internet services on demand. Prominent service providers like Amazon, Google, SUN, IBM, Oracle, Salesforce etc. are extending computing infrastructures and platforms as a core for providing top-level services for computation, storage, database and applications. Application services could be email, office applications, finance, video, audio and data processing.

During our analysis, we counted more than 70 providers of so called Cloud services and a selected number of them are summarized in Tab. ***1***.

Tab. 1 Offerings of Services On-Demand

Company/Product	Service type	Pricing Model	CBMF Concept
Amazon EC2, S3, SimpleDB, SQS, FPS, DevPay	Computing, Storage, Database, Payment, Billing	Pay-per-use	Infrastructure/ Platform-as-a-Service
Appian Anywhere	Business Process Management	Pay-per-use	Applications
Box.net	Storage	Pay-per-use	Applications
FlexiScale	Infrastructure	Pay-per-use	Infrastructure
Google App Engine	Infrastructure, Web Applications	Pay-per-use	Infrastructure

[24] http://www.mor.ph/
[25] http://appengine.google.com/
[26] http://www.bungeelabs.com/platform/
[27] http://www.salesforce.com/platform/
[28] http://www.google.com/a/
[29] http://www.sap.com/solutions/sme/businessbydesign/index.epx
[30] http://www.xignite.com
[31] http://www.strikeiron.com

Gmail Drive	Storage, Email	free* / Pay-per-use	Applications
MuxCloud	Data Processing (Video): uses Amazon’s EC2	Pay-per-use	Applications
Nirvanix	Storage	Pay-per-use	Applications
Network.com	Infrastructure	Pay-per-use	Infrastructure/ Platform-as-a-Service
OpSource	Billing	Subscription	Applications
Process Maker Live	Business Process Management	Pay-per-use	Applications
Salesforce.com	Platform	Pay-per-use	Platform-as-a-Service/ Applications
MS SkyDrive	Storage	free*	Applications
SmugMug	Data Sharing (Photo)	Subscription	Applications
Strikeiron	Web Services	Subscription/ Pay-per-use	Applications
XDrive	Storage	Subscription	Applications
XCalibre	Infrastructure	Subscription	Infrastructure
Zimory.com	Marketplace	Dynamic pricing	Applications

* free until some limited contingent e.g. 5GB or 7GB

The table gives an overview of the providers, the service types they are offering, the selected pricing models as well as a mapping to a concept of the proposed framework.

The offered services are categorized in the service types *Infrastructure*, *Storage*, *Database*, *Business Process Management*, *Marketplace*, *Billing*, *Accounting*, *Email*, *Data sharing*, *Data processing* and *Web services*.

The most used pricing model of the presented service providers is *Pay-per-use*, where the user pays a static price for a used unit, often per hour, GB, CPU-hour etc. The *Pay-per-use* pricing model is a simple model, where units (or units per time) are associated with fixed price values. This simple concept is widely used for products (services), where mass production and widespread delivery made price negotiation impractical (Wurman 2001, Bitran and Caldentey 2003). A similar, but different pricing model is *Subscription*, where the user subscribes (signs a contract) for using a pre-selected combination of service units for a fixed price and longer time frame, usually monthly or yearly.

The dominance of the above pricing models can be explained due to the fact, that users often prefer simple pricing models (like *Pay-per-use* or *Subscription*), where the payment fee is static. The reasons might be the ease of understanding and accounting, the clear base of calculation but also psychological reasons, such as overestimation of usage and avoidance of occasional large bills, even when the fixed-fee costs more over time (Fishburn and Odlyzko 1999).

Dynamic pricing (called also variable pricing) is a pricing model, where the target service price is established as a result of dynamic supply and demand, e.g. by means of auctions or negotiations. This pricing model is typically used for calculating the price of differentiated and high value items. Auctions are standard mechanisms for performing aggregations of supply and demand (Wurman 2001). Lai (2005) illustrates that market mechanisms, implementing dynamic pricing policies could achieve better economically efficient allocations and prices of differentiated high-value services.

In a market, where Cloud providers face scarce resources and thus high demand, capacity allocation depends on customer choice, customer classification and appropriate pricing. Business and decision concepts like in Revenue Management are paving the way for a successful application of a well-known decision framework for Cloud providers. A precise selection of consumers will gain higher revenue (Anandasivam and Neumann 2009).

4 Conclusions and New Research Directions

Cloud Computing and related computing paradigms and concepts like Grid Computing, Utility Computing or Voluntary Computing have been controversially discussed in academia and industry. This paper focused on giving a clear distinction between Cloud Computing and Grid Computing by identifying a catalogue of criteria and comparing both paradigms.

In order to propose business models for Cloud Computing that ensure the sustainability of these systems, a Cloud Business Model Framework was introduced built of three layers: infrastructure, platforms and applications in Clouds. Several providers were analyzed and categorized by their services and pricing models based on this framework.

Subsequently, this paper will now discuss new research areas in the context of Cloud Computing. These issues leave ample room for both technical and economic future work to meet the requirements of services in Clouds and make the Cloud vision come true.

4.1 Application Domains

In this section, just a few but maybe typical application domains should be named: Nowadays, servers with high performance capabilities are necessary nearly across all industry sectors. Amazon offers different kinds of operating systems with preconfigured settings (so called Amazon Machine Instance or AMI) for instant usage on the Infrastructure as a Service (IaaS) level. For example, the Havard Medical School runs an AMI with a customized Oracle Database for genetic testing purposes. The virtualization technology enables Amazon to preconfigure a huge amount of specific VMs. Amazon benefits from offering these preconfigured machines to a wide range of customers to satisfy different needs.

Beyond that, using IaaS for running tests of huge information systems such as (modules of) ERP-Systems or other complex planning systems is very effective. Instead of maintaining the capacities for all different testing purposes in a company, the dynamic purchase/procurement of resources obviously is much more cost effective and technically flexible. Another important application area are all interactive web applications for a growing number of customers, particularly if the number might increase at any (unpredictable) point of time – IaaS provides a scalable technology for supporting all kinds of this scenario.

4.2 Cloud API

In order to allow the easy use of Cloud systems and to enable the migration of applications between the Clouds of different providers, there will be the need for a standardized Cloud API. While there are first approaches (e.g. Enomaly RESTful API[32]), there is no commonly agreed-upon API or Cloud reference implementation that programmers can rely on. At this point, it might be too early for such a reference implementation. It might be beneficial to have multiple competing architectures and approaches in order to be able to single out best practices. However, a standard will be required in the long run in order to make the vision of the

[32] https://enomalism.svn.sourceforge.net/svnroot/enomalism/enomalism_rest_api.pdf

Cloud come true. It might last until the Cloud market has settled down and is consolidated to few large players, each with distinct capabilities and target segments. Until then, Cloud providers will try to lock in their customers with proprietary interfaces.

4.3 Business Models

Clouds also require new business models, especially with respect to the licensing of software. This has already become an issue in Grids. Current licenses are bound to a single user or physical machine. However, in order to allow the efficient sharing of resources and software licenses, there has to be a move towards usage-based licenses, penalties and pricing (Becker et al. 2008), very much in the spirit of Utility Computing. Software vendors have to realize that Clouds will especially enable and drive the adoption of Software-as-a-Service models.

4.4 Pricing of Complex Services

A tremendous increase of service offerings especially in the Cloud Computing area and an emergence of more sophisticated enabler technologies for service composition and ad-hoc creation of situational applications are observable. These approaches attracted attention in the business context as they increase flexibility, lower fix costs and consequently eliminate lock-in situations. A commercial success can only be achieved by developing adequate pricing models that foster an efficient way to allocate and valuate composite services. Blau et al. (2008) present a multidimensional procurement auction for composite services. They provide a model for service value networks based on a graph structure. The auction mechanism allocates a path through this network that represents a technical feasible composite service based on the service providers' bids containing price and configuration of their offerings. The mechanism design incentivizes participants to reveal their true valuation for their service offering as well as its true configuration.

4.5 The Long Tail in Clouds

The hype about Cloud computing accompanies the Software as a Service (SaaS) wave. Services built on top of Cloud infrastructures enable software providers to offer products at lower cost and simultaneously with a higher degree of customization. This so-called “Long Tail” departs from the mass market and focuses on many niche markets (Anderson 2008). Cloud Computing enables the access to large data centers for Small and Medium Enterprises (SME), which they can use to provide unique services on large-scale resources.

However, the selling of few unique services requires a thorough understanding of portfolio management. Unlike mass management, the Long Tail strategy needs a continuous improvement and change of the currently offered products. Both scientific design methods and analytical tools have to be developed and applied for services in Clouds to identify the specific characteristics of the Cloud environment. An interdisciplinary approach taking microeconomy, technological feasibility and business models into account will provide insights into this complex interaction (Weinhardt et al. 2003).

4.6 Security and Trust

Another crucial point for the eventual acceptance of Cloud technology in business industries will be the safety of critical data, both in transfer as in storage. Large enterprises will not be willing to support the Cloud concept as long as there is not more transparency available at which geographical location the data is stored and how it is protected (Henschen 2008). Reasons for that are foreign laws, which would possibly allow foreign governments to access this data, or domestic insurance contracts, demanding the data to be stored only in certain regions. But providing this required, transparency would contradict in some ways the whole idea of Cloud Computing itself, and it remains to be seen how the big Cloud vendors will tackle this concern.

Bibliography

Anandasivam A, Neumann D (2009) Managing Revenue in Grids. In: Proceedings of the 42nd Hawaii International Conference on System Sciences

Anderson C (2008) The Long Tail. Hyperion, New York, USA

Becker M, Borissov N, Deora V, Rana O, Neumann D (2008) Using k-pricing for Penalty Calculation in Grid Markets. In: Proceedings of the 41st Hawaii International Conference on System Sciences

Bitran G, Caldentey R (2003) An Overview of Pricing Models for Revenue Management. Manufacturing & Service Operations Management 5(3):203-229

Blau B, Neumann D, Weinhardt C, Michalk W (2008) Provisioning of Service Mashup Topologies. In: Proceedings of the 16th European Conference on Information Systems

Boss G, Malladi P, Quan S, Legregni L, Hall H (2007) Cloud Computing. Technical Report, IBM High Performance On Demand Solutions, 8 October 2007, Version 1.0

EGEE (2008) An EGEE Comparative Study: Grids and Clouds – Evolution or Revolution. https://edms.cern.ch/document/925013. Last accessed 2008-06-25

Fishburn PC, Odlyzko AM (1999) Competitive Pricing of Information Goods: Subscription Pricing Versus Pay-Per-Use. Economic Theory 3: 447-470

Foster I, Kesselman C (1999) The Grid: Blueprint for a Future Computing Infrastructure. Morgan Kaufmann, San Mateo, CA

Foster I, Kesselman C, Tuecke S (2001) The Anatomy of the Grid – Enabling Scalable Virtual Organizations. International Journal of High Performance Computing Applications 15(3):200-222

Foster I (2002) What is the Grid? A Three Point Checklist. GRIDToday 1(6):22-25

Gentzsch W (2008) Grids are Dead! Or are They? http://www.on-demandenterprise.com/features/26060699.html. Last accessed 2009-03-30

Henschen D (2008) Demystifying Cloud Computing. http://www.intelligententerprise.com/blog/archives/2008/06/demystifying_cl.html. Last accessed 2009-03-30

IBM (2006) Grid Computing: Past, Present and Future - An Innovation Perspective. Technical Report. http://www-1.ibm.com/grid/pdf/innovperspective.pdf. Last accessed 2009-03-30

Lai K (2005) Markets are Dead, Long live Markets. SIGecom Exchanges 5(4):1-10

Lawton G (2008) Moving the OS to the Web. Computer 41(3):16-19

Neumann D, Stößer J, Weinhardt C, Nimis J (2008) A Framework for Commercial Grids - Economic and Technical Challenges. Journal of Grid Computing 6(3):325-347

Rappa M (2004) The Utility Business Model and the Future of Computing Services. IBM Systems Journal 43(1):32-42

Skillicorn D (2002) The Case for Datacentric Grids. In: Proceedings of the International Parallel and Distributed Processing Symposium

Weinhardt C, Holtmann C, Neumann D (2003) Market-Engineering. Wirtschaftsinformatik 45(6):635-640

Weiss A (2007) Computing in the Clouds. netWorker 11(4):16-25

Wurman PR (2001) Dynamic Pricing in the Virtual Marketplace. IEEE Internet Computing 5(2):36-42