The Future Of Cloud Computing

Opportunities For European Cloud Computing Beyond 2010

••• Expert Group Report

Public Version 1.0

Rapporteur for this Report: Lutz Schubert [USTUTT-HLRS]

Editors: Keith Jeffery [ERCIM], Burkhard Neidecker-Lutz [SAP Research]

European Commission
Information Society and Media

EXECUTIVE SUMMARY

Though the concept of "clouds" is not new, it is undisputable that they have proven a major commercial success over recent years and will play a large part in the ICT domain over the next 10 years or more, as future systems will exploit the capabilities of managed services and resource provisioning further. Clouds are of particular commercial interest not only with the growing tendency to outsource IT so as to reduce management overhead and to extend existing, limited IT infrastructures, but even more importantly, they reduce the entrance barrier for new service providers to offer their respective capabilities to a wide market with a minimum of entry costs and infrastructure requirements – in fact, the special capabilities of cloud infrastructures allow providers to experiment with novel service types whilst reducing the risk of wasting resources.

Cloud systems are not to be misunderstood as just another form of resource provisioning infrastructure and in fact, as this report shows, multiple opportunities arise from the principles for cloud infrastructures that will enable further types of applications, reduced development and provisioning time of different services. Cloud computing has particular characteristics that distinguish it from classical resource and service provisioning environments:

(1) it is (more-or-less) infinitely scalable; (2) it provides one or more of an infrastructure for platforms, a platform for applications or applications (via services) themselves; (3) thus clouds can be used for every purpose from disaster recovery/business continuity through to a fully outsourced ICT service for an organisation; (4) clouds shift the costs for a business opportunity from CAPEX to OPEX which allows finer control of expenditure and avoids costly asset acquisition and maintenance reducing the entry threshold barrier; (5) currently the major cloud providers had already invested in large scale infrastructure and now offer a cloud service to exploit it; (6) as a consequence the cloud offerings are heterogeneous and without agreed interfaces; (7) cloud providers essentially provide datacentres for outsourcing; (8) there are concerns over security if a business places its valuable knowledge, information and data on an external service; (9) there are concerns over availability and business continuity – with some recent examples of failures; (10) there are concerns over data shipping over anticipated broadband speeds.

The concept of cloud computing is linked intimately with those of IaaS (Infrastructure as a Service); PaaS (Platform as a Service), SaaS (Software as a Service) and collectively *aaS (Everything as a Service) all of which imply a service-oriented architecture.

Open Research Issues

CLOUD TECHNOLOGIES AND MODELS HAVE NOT YET REACHED THEIR FULL POTENTIAL AND MANY OF THE CAPABILITIES ASSOCIATED WITH CLOUDS ARE NOT YET DEVELOPED AND RESEARCHED TO A DEGREE THAT ALLOWS THEIR EXPLOITATION TO THE FULL DEGREE, RESPECTIVELY MEETING ALL REQUIREMENTS UNDER ALL POTENTIAL CIRCUMSTANCES OF USAGE.

Many aspects are still in an experimental stage where the long-term impact on provisioning and usage is as yet unknown. Furthermore, plenty of as yet unforeseen challenges arise from exploiting the cloud capabilities to their full potential, involving in particular aspects deriving from the large degree of scalability and heterogeneity of the underlying resources. We can thereby distinguish between *technological* gaps on the one hand, that need to be closed in order to realize cloud infrastructures that fulfil the specific cloud characteristics and *non-technological* issues on the other hand that in particular reduce uptake and viability of cloud systems:

To the *technological* aspects belong in particular issues related to (1) scale and elastic scalability, which is not only currently restricted to horizontal scale out, but also inefficient as it tends to resource over usage due to limited scale down capabilities and full replication of instances rather than only of essential segments. (2) Trust, security and privacy always pose issues in any internet provided service, but due to the specific nature of clouds, additional aspects related e.g. to multi-tenancy arise and control over data location etc. arise. What is more, clouds simplify malicious use of resources, e.g. for hacking purposes, but also for sensitive calculations (such as weapon design) etc. (3) Handling data in clouds is still complicated - in particular as data size and diversity grows, pure replication is no viable approach, leading to consistency and efficiency issues. Also, the lacking control over data location and missing provenance poses security and legalistic issues. (4) Programming models are currently not aligned to highly scalable applications and thus do not exploit the capabilities of clouds, whilst they should also simplify development. Along the same line, developers, providers and users should be able to control and restrict distribution and scaling behaviour. This relates to (5) systems development and management which is currently still executed mostly manually, thus contributing to substantial efficiency and bottleneck issues.

On the other hand, *non-technological* issues play a major role in realizing these technological aspects and in ensuring viability of the infrastructures in the first instance. To these belong in particular (1) economic aspects which cover knowledge about when, why, how to use which cloud system how this impacts on the original infrastructure (provider) –long-term experience is lacking in all these areas; and (2) legalistic issues which come as a consequence from the dynamic (location) handling of the clouds, their scalability and the partially unclear legislative issues in the internet. This covers in particular issues related to intellectual property rights and data protection. In addition, (3) aspects related to green IT need to be elaborated further, as the cloud offers principally "green capabilities" by reducing unnecessary power consumption, given that good scaling behaviour and good economic models are in place.

Europe and Clouds

Notwithstanding common beliefs, clouds are not a phenomenon entirely imported from abroad. This report will elaborate the main opportunities for European industry *and* research to be pursued with respect to the specific capabilities and remaining gaps.

This document provides a detailed analysis of Europe's position with respect to cloud provisioning, and how this affects in particular future research and development in this area. The report is based on a series of workshops involving experts from different areas related to cloud technologies.

EUROPE'S MAIN OPPORTUNITIES TO PARTICIPATE IN THE "CLOUD MOVEMENT" CONSIST IN PARTICULAR IN ASPECTS RELATED TO EXTENDING AND COMPLETING THE CAPABILITIES OF CURRENT CLOUD SYSTEMS, WHEREBY THE LONG-TERM GOAL CONSISTS IN REALIZING META-SCALABLE CLOUD SYSTEMS AND SERVICES. THE COMPLEXITY TO REALIZE THE OPPORTUNITIES DIRECTLY DEPENDS ON THE COMPLEXITY TO PERFORM THE UNDERLYING RESEARCH WORK AND OF THE CURRENT DEVELOPMENT STATUS.

In more detail, the identified opportunities are: (1) Provisioning and further development of Cloud infrastructures, where in particular telecommunication companies are expected to provide offerings; (2) Provisioning and advancing cloud platforms, which the telecommunication industry might see as a business opportunity, as well as large IT companies with business in Europe and even large non-IT businesses with hardware not fully utilised. (3) Enhanced service provisioning and development of meta-services: Europe could and should develop a 'free market for IT services' to match those for movement of goods, services, capital, and skills. Again telecommunication industry could supplement their services as ISPs with extended cloud capabilities; (4) provision of

consultancy to assist businesses to migrate to, and utilise effectively, clouds. This implies also provision of a toolset to assist in analysis and migration.

Recommendations Overview

Due to the strong commercial nature of cloud systems, both technological and non-technological aspects are involved in cloud provisioning. Since both areas still have major gaps, the recommendations are not restricted to purely technological issues, but also cover non-technological aspects related in particular to the economical and legalistic side of cloud systems.

Europe is in a strong position to address both these areas: technologically due to its excellent background in many of the key research and development aspects related to cloud systems, such as GRIDs and Service Oriented Architectures, and non-technologically due to Europe's position as a united body. Europe also has a strong market position with many of major contributors from different field originate from Europe.

The recommendations towards research and development communities and bodies as expressed in this report hence do address a wide scope of outstanding issues, ranging from specific research and development topics over general policies to legalistic aspects which currently pose a major obstacle towards wide uptake of cloud infrastructures:

Main Recommendations

Recommendation 1: The EC should stimulate research and technological development in the area of Cloud Computing

	Cloud computing poses a variety of challenges to conventional advanced ICT, mostly due to the fact of the unprecedented scale and heterogeneity of the required infrastructure. This demands a rethinking of even current advanced ICT solutions. Plenty of research issues remain to be addressed in the context of cloud provisioning. Europe should exploit the available expertise and results from areas such as Grid, Service Oriented Architectures and e-infrastructure to help realizing the next generation of services on cloud systems. Particular research topics to be addressed further are: (1) Elastic scalability; (2) Cloud (systems) development and management; (3) Data management; (4) Programming models and resource control; (5) trust, security and privacy.

Recommendation 2: The EC together with Member States should set up the right regulatory framework to facilitate the uptake of Cloud computing

	Cloud systems are mostly in an experimental stage – to fully exploit their capabilities in particular from a commercial side, the according impact, dependencies, requirements etc. need to be evaluated carefully. Accordingly, research efforts need to be vested not only into technological aspects of *realizing* cloud systems, but also into the aspects related to commercial and business aspects, in particular involving economical and legalistic concerns. Accordingly, business consultants, legal researchers, governmental bodies etc. should be encouraged to participate in investigating the particular circumstances of cloud provisioning. Obviously, technologies thereby need to recognize results from these areas, just as economical and legalistic views need to acknowledge the technological capabilities and restrictions. In summary, the specific issues are: (1) Economical aspects; (2) Legalistic issues; (3) Green IT.

Additional Recommendations

Additional Recommendation 1: The EU needs large scale research and experimentation test beds

	A major obstacle for European research communities to develop and test effective large scale cloud systems consists in the lack of available resource infrastructures of a size that allow experimentation and testing. Such infrastructure test beds could be provided through joint, collaborative efforts between existing resource owners and public, as well as non-public research bodies, e.g. through public-private partnerships or through fostering existing research communities building up on public e-infrastructures etc.

Additional Recommendation 2: The EC together with industrial and public stakeholders should develop joint programmes encourage expert collaboration groups

	The development of future cloud infrastructures and in particular of meta-clouds necessitates the collaboration of experts from various backgrounds related to cloud systems, as can be implicitly seen from the recommendations above. This would include research and development, academia and industry equally. To encourage such collaboration, the need for interoperable meta-clouds needs to be promoted more clearly.

Additional Recommendation 3: The EC should encourage the development and production of (a) CLOUD interoperation standards (b) an open source reference implementation

	The development of standards and a reference implementation would assist European SMEs in particular in ensuring their products and service offerings in the cloud environment have the widest possible market and customer acceptability. The standards should encourage all suppliers to be able to interoperate; the reference implementation is to allow plug-tests to prove standards compliance.

Additional Recommendation 4: The EC should promote the European leadership position in software through commercially relevant open source approaches

	Maintaining an open source approach for research results and cloud infrastructure support tools ensures uptake and simplifies adaptation to different environments. The European open source movement should thereby work strongly together with industry to support commercial cloud based service provisioning.

I. THE ADVENT OF THE "CLOUDS"

The increased degree of connectivity and the increasing amount of data has led many providers and in particular data centres to employ larger infrastructures with dynamic load and access balancing. By distributing and replicating data across servers on demand, resource utilisation has been significantly improved. Similarly web server hosts replicate images of relevant customers who requested a certain degree of accessibility across multiple servers and route requests according to traffic load.

However, it was only when Amazon published these internal resources and their management mechanisms for use by customers that the term "cloud" was publicly associated with such elastic infrastructures – especially with "on demand" access to IT resources in mind. In the meantime, many providers have rebranded their infrastructures to "clouds", even though this had little consequences on the way they provided their capabilities.

It may be noted in this context that the term "cloud" dates back to the 90s in reference to the capability of dynamic traffic switching to balance utilization ("telecom clouds") and to indicate that the telecoms infrastructure is virtualised – the end user does not know or care over which channels her data is routed (see IETF meeting minutes [1]). Microsoft adopted the term 2001 in a public presentation about the .NET framework to refer to the infrastructure of computers that make up the internet [2]. According to Wikipedia, the underlying concept of cloud computing can be dated even further back to a public speech given by John McCarthy 1961 where he predicts that computer time-sharing may lead to the provisioning of computing resources and applications as a utility [3].

Concept and even technological approaches behind "cloud computing" can thus not be considered a novelty as such and in particular data centres already employed methods to maintain scalability and reliability to ensure availability of their hosted data. What is more, cloud systems are, unlike e.g. grid computing, not driven by research first and then being taken up by industry, but instead originates directly from commercial requirements and solutions. It is hence not surprising, that the term "cloud computing" and its current understanding only really became popular with Amazon's publication of the Elastic Compute Cloud EC2 in 2006 [4], giving rise to a small boom of "cloud offerings" which mostly consisted in a rebranding of their existent in-house solutions and techniques, as well as a potential exposition of these capabilities to consumers.

Multiple new "cloud" domains and providers have thus arisen and it is not surprising, that the term has found multiple related, yet different meanings. In particular, the scope of areas and capabilities that so-called clouds are applied for differs thereby strongly. The most typical representatives for cloud related functionalities can currently be found in the following areas: (1) data centres trying to maintain high scalability and increase availability; (2) web server farms automating and stabilising their servers, respectively the user's website; (3) in house attempts to balance resources over the business solutions; (4) external ASP-type offerings.

It must be made clear in this context that "Clouds" do generally not refer to a specific technology or framework, but rather to a set of combined technologies, respectively a paradigm / concept. The "Grid" and Service Oriented Architectures are often confused as being identical with clouds due to this primarily conceptual understanding (see also section II.C). Likewise, current "cloud providers" typically build upon proprietary technology sets and approaches based on their in-house solutions - only little efforts have been undertaken so far, to build up a generic framework / middleware supporting all the features related to clouds.

It's only been in 2004 that multi-core processing became available for common desktop machines, when Intel finally abandoned the development of a 4 GHz processor and switched to multi-core development instead [5]. Implicitly even more mainstream developers and users investigate the specific advantages and problems of not only horizontal, but also vertical scalability. Additionally, with the "Prosumer" [6] movement, as well as the growing demand to lower management cost and the carbon footprint make outsourcing more and more interesting for the market.

It is to be expected that the cloud paradigm will find further uptake in the future – not only as a means to manage the infrastructure of providers, but also to provide smaller entities with the capabilities of a larger infrastructure that they cannot afford to own themselves. At the same time, the cloud paradigm will allow for a set of enhanced capabilities and services not feasible before.

1. Clouds in the Future Internet

The Future Internet covers all research and development activities dedicated to realizing tomorrow's internet, i.e. enhancing a networking infrastructure which integrates all kind of resources, usage domains etc. As such, research related to cloud technologies form a vital part of the Future Internet research agenda. Confusions regarding the aspects covered by cloud computing with respect to the Future Internet mostly arise from the broad scope of characteristics assigned to "clouds", as is the logical consequence of the re-branding boom some years ago.

So far, most cloud systems have focused on hosting applications and data on remote computers, employing in particular replication strategies to ensure availability and thus achieving a load-balancing scalability. However, the conceptual model of clouds exceeds such a simple technical approach and leads to challenges not unlike the ones of the future internet, yet with slightly different focus due to the combination of concepts and goals implicit to cloud systems.

In other words, as a technological realisation driven by an economic proposition, cloud infrastructures would offer capabilities that enable relevant aspects of the future internet, in particular related to scalability, reliability and adaptability. At the same time, the cloud concept addresses multiple facets of these functionalities.

A. About This Report

This report was initiated by the European Commission in 2009 to capture the development in cloud computing and its relevance and meaning for the European market and research communities. It bases on a series of meetings between invited experts that discussed the current technological and economic situation, its development in the near and far future, as well as future requirements towards cloud technologies to enable and maximize a European economic opportunity.

Cloud computing is a huge field as such and the impact on and relevance for Europe is difficult to capture. Cloud technologies are evolving already and the current development runs a high risk of ending in proprietary solutions which only cover aspects of the overall concept. The present report tries to bring together the individual experts' perspectives and highlights the main issues considered relevant in the future.

Document Structure

The document is structured into 5 main sections (and two appendixes), following the main analysis process:

Chapter I provides some background information about the report and history of cloud system, thus providing the context of this document.

Chapter II elaborates the different concepts related to cloud systems: being mostly a marketing term, "cloud" is used differently in various contexts. This chapter explains how the terms and concepts are applied in the context of this report and also positions clouds with respect to other related areas that are often confused with cloud systems. Appendix A will extend this discussion with areas that may play a long-term impact on cloud infrastructures.

Chapter III analyses the current state of the art from the perspective of both commercial development and (academic) research with a particular focus on identifying the open issues with respect to the specific capabilities associated / requested from clouds. Whilst the chapter does not claim to provide a complete, exhaustive state of the art analysis, it does capture the essence of what users and uptakers can expect from current and near-future technologies in this domain.

Chapter IV performs a detailed analysis of the European position in the cloud movement, its strengths, weaknesses, threats and in particular the specific opportunities where the European research communities and industrial players could and should contribute in the realization of future cloud systems. Basing on the gaps identified in chapter III, this chapter also provides a quick overview over the main areas of potential interest for European RTD given its specific strengths. Specific use case scenarios of future cloud systems will also be further elaborated in Appendix B.

Chapter V concludes the analysis of this report with an in-depth examination of the gaps (chapter III) and opportunities (chapter IV) to identify the specific recommendations that can be made for European research and development. In particular it identifies the dependencies between research and development topics towards realization of the specific opportunities for Europe.

B. Acknowledgments

This report was made possible by the European Commission; particular thanks go to Maria Tsakali, Jesús Villasante, David Callahan, Arian Zwegers and Jorge Gasós for organization of the workshops and contribution to the report.

The report could not have been realized without the excellent contributions from all workshop participants.

C. List of Experts

Prashant Barot [Oracle], Francis Behr [Syntec], Peter Bosch [Alcatel Lucent], Ivona Brandic [Vienna University of Technology], Brigitte Cardinael [France Telecom], Thierry Coupaye [France Telecom Orange Labs], Richard Davies [Elastichosts], David De Roure [University of Southampton], Philippe Dobbelaere [Alcatel Lucent], Andreas Ebert [Microsoft], Aake Edlund [KTH], Guido Falkenberg [Software AG], Jürgen Falkner [Fraunhofer], William Fellows [The 451 Group], Friedrich Ferstl [SUN], Ioannis Fikouras [Ericsson], Mike Fisher [British Telecom], Behrend Freese [Zimory], Alfred Geiger [T-Systems], Juanjo Hierro [Telefonica I&D], Giles Hoghen [ENISA], Keith Jeffery [ERCIM], Ricardo Jimenez-Peris [UPM], Ruby Krishnaswamy [France Telecom Orange Labs], Frank Leymann [University of Stuttgart - IAAS], Ignacio Llorente [UCM], Monica Marinucci [Oracle], Joan Masso [Gridsystems], Cyril Meunier [IDC], Christine Morin [INRIA], Sebastian Müller [Google], Burkhard Neidecker-Lutz [SAP], Mathieu Poujol [PAC], Thierry Priol [INRIA], Harald Schöning [Software AG], Lutz Schubert [High Performance Computing Center Stuttgart], Dave Snelling [Fujitsu Labs Europe], Paul Strong [eBay], Werner Teppe [Amadeus], Clemems Thole [Fraunhofer], Dora Varvarigou [NTUA], Stefan Wesner [High Performance Computing Center Stuttgart], Per Willars [Ericsson], Yaron Wolfsthal [IBM], Hans Wortmann [University of Groningen].

II. WHAT IS A "CLOUD"

Various definitions and interpretations of "clouds" and / or "cloud computing" exist. With particular respect to the various usage scopes the term is employed to, we will try to give a *representative* (as opposed to complete) set of definitions as recommendation towards future usage in the cloud computing related research space. This report does not claim completeness with this respect, as it does not introduce a new terminology, but tries to capture an abstract term in a way that best represents the technological aspects and issues related to it.

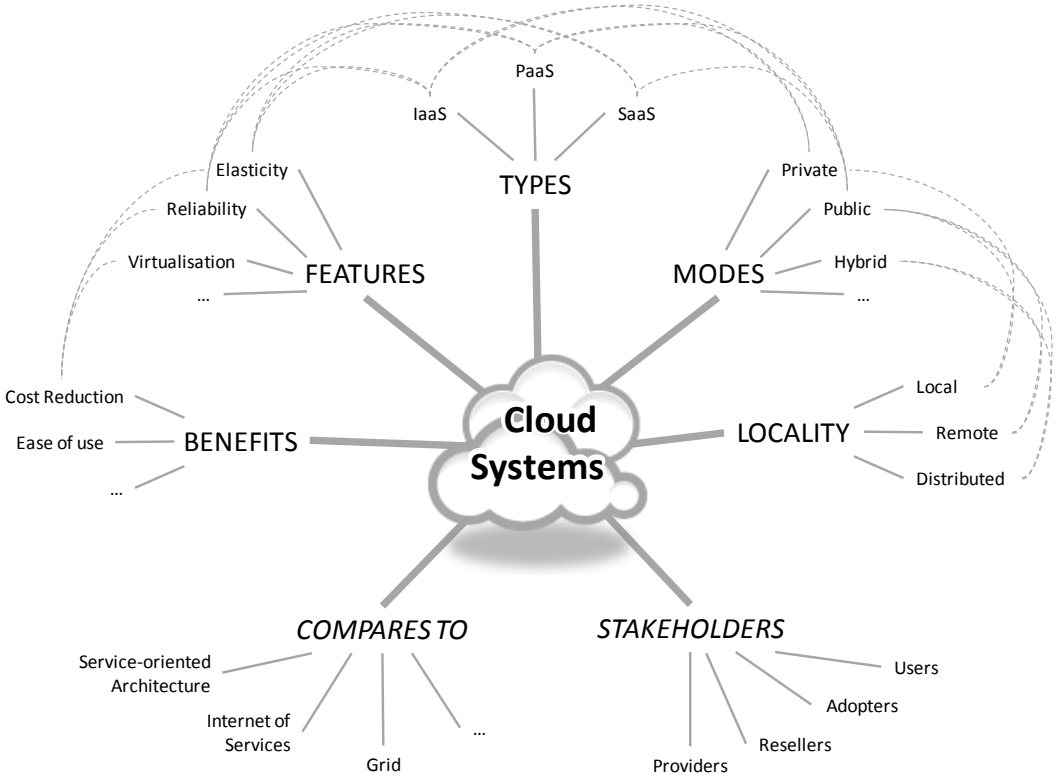

FIGURE 1: NON-EXHAUSTIVE VIEW ON THE MAIN ASPECTS FORMING A CLOUD SYSTEM

A. TERMINOLOGY

In its broadest form, we can define

> *a 'cloud' is an elastic execution environment of resources involving multiple stakeholders and providing a metered service at multiple granularities for a specified level of quality (of service).*

In other words, clouds as we understand them in the context of this document are primarily platforms that allow execution in various forms (see below) across multiple resources (and potentially across enterprise boundaries, see below) – the main characteristics will be detailed in section II.B. We can distinguish different *types* of clouds (cf. section II.A.1), all of which have in common that they (directly or indirectly) enhance resources and services with additional capabilities related to manageability, elasticity and system platform independency.

To be more specific, a cloud is a platform or infrastructure that enables execution of code (services, applications etc.), in a managed and elastic fashion, whereas "managed" means that reliability according to pre-defined quality parameters is automatically ensured and "elastic" implies that the

resources are put to use according to actual current requirements observing overarching requirement definitions – implicitly, elasticity includes both up- and downward scalability of resources and data, but also load-balancing of data throughput.

As shall be elaborated, future cloud systems should also be able to maintain a pre-specified level of quality, respectively boundary conditions (including performance, energy consumption, etc.) and should allow integration of resources across organisational boundaries, integrating multiple stakeholders.

Noticeably, the actual details of the capabilities differ slightly depending on how the cloud is employed: since clouds relate to a usage concept, rather than a technology, it has been applied to different areas, as described in the introductory part of this document. We therefore need to distinguish what kinds of capabilities are provided by different cloud systems:

1. Types of Clouds

Cloud providers typically centre on one type of cloud functionality provisioning: Infrastructure, Platform or Software / Application, though there is potentially no restriction to offer multiple types at the same time, which can often be observed in PaaS (Platform as a Service) providers which offer specific applications too, such as Google App Engine in combination with Google Docs. Due this combinatorial capability, these types are also often referred to as "components" (see e.g. [7]).

Literature and publications typically differ slightly in the terminologies applied. This is mostly due to the fact that some application areas overlap and are therefore difficult to distinguish. As an example, platforms typically have to provide access to resources indirectly, and thus are sometimes confused with infrastructures. Additionally, more popular terms have been introduced in less technologically centred publications.

The following list identifies the main types of clouds (currently in use):

(Cloud) Infrastructure as a Service (IaaS) also referred to as *Resource Clouds*, provide (managed and scalable) resources as services to the user – in other words, they basically provide enhanced virtualisation capabilities. Accordingly, different resources may be provided via a service interface:

Data & Storage Clouds deal with reliable access to data of potentially dynamic size, weighing resource usage with access requirements and / or quality definition.

Examples: Amazon S3, SQL Azure.

Compute Clouds provide access to computational resources, i.e. CPUs. So far, such low-level resources cannot really be exploited on their own, so that they are typically exposed as part of a "virtualized environment" (not to be mixed with PaaS below), i.e. hypervisors. Compute Cloud Providers therefore typically offer the capability to provide computing resources (i.e. raw access to resources unlike PaaS that offer full software stacks to develop and build applications), typically virtualised, in which to execute cloudified services and applications. IaaS (Infrastructure as a Service) offers additional capabilities over a simple compute service.

Examples: Amazon EC2, Zimory, Elastichosts.

(Cloud) Platform as a Service (PaaS), provide computational resources via a *platform* upon which applications and services can be developed and hosted. PaaS typically makes use of dedicated APIs to control the behaviour of a server hosting engine which executes and replicates the execution according to user requests (e.g. access rate). As each provider exposes his / her own API according

to the respective key capabilities, applications developed for one specific cloud provider cannot be moved to another cloud host – there are however attempts to extend generic programming models with cloud capabilities (such as MS Azure).

Examples: Force.com, Google App Engine, Windows Azure (Platform).

(Clouds) Software as a Service (SaaS), also sometimes referred to as *Service or Application Clouds* are offering implementations of specific business functions and business processes that are provided with specific cloud capabilities, i.e. they provide applications / services *using* a cloud infrastructure or platform, rather than providing cloud features themselves. Often, kind of standard application software functionality is offered within a cloud.

Examples: Google Docs, Salesforce CRM, SAP Business by Design.

Overall, Cloud Computing is not restricted to Infrastructure / Platform / Software as a Service systems, even though it provides enhanced capabilities which act as (vertical) enablers to these systems. As such, I/P/SaaS can be considered specific "usage patterns" for cloud systems which relate to models already approached by Grid, Web Services etc. Cloud systems are a promising way to implement these models and extend them further.

2. Deployment Types (Cloud Usage)

Similar to P/I/SaaS, clouds may be hosted and employed in different fashions, depending on the use case, respectively the business model of the provider. So far, there has been a tendency of clouds to evolve from private, internal solutions (private clouds) to manage the local infrastructure and the amount of requests e.g. to ensure availability of highly requested data. This is due to the fact that data centres initiating cloud capabilities made use of these features for internal purposes before considering selling the capabilities publicly (public clouds). Only now that the providers have gained confidence in publication and exposition of cloud features do the first hybrid solutions emerge. This movement from private via public to combined solutions is often considered a "natural" evolution of such systems, though there is no reason for providers to not start up with hybrid solutions, once the necessary technologies have reached a mature enough position.

We can hence distinguish between the following deployment types:

Private Clouds are typically owned by the respective enterprise and / or leased. Functionalities are not directly exposed to the customer, though in some cases services with cloud enhanced features may be offered – this is similar to (Cloud) Software as a Service from the customer point of view.

Example: eBay.

Public Clouds. Enterprises may use cloud functionality from others, respectively offer their own services to users outside of the company. Providing the user with the actual capability to exploit the cloud features for his / her own purposes also allows other enterprises to outsource their services to such cloud providers, thus reducing costs and effort to build up their own infrastructure. As noted in the context of cloud *types*, the scope of functionalities thereby may differ.

Example: Amazon, Google Apps, Windows Azure.

Hybrid Clouds. Though public clouds allow enterprises to outsource parts of their infrastructure to cloud providers, they at the same time would lose control over the resources and the distribution / management of code and data. In some cases, this is not desired by the respective enterprise. *Hybrid clouds* consist of a mixed employment of *private* and *public cloud* infrastructures so as to

achieve a maximum of cost reduction through outsourcing whilst maintaining the desired degree of control over e.g. sensitive data by employing local private clouds.

There are not many hybrid clouds actually in use today, though initial initiatives such as the one by IBM and Juniper already introduce base technologies for their realization [11].

Community Clouds. Typically cloud systems are restricted to the local infrastructure, i.e. providers of public clouds offer their own infrastructure to customers. Though the provider could actually resell the infrastructure of another provider, clouds do not *aggregate* infrastructures to build up larger, cross-boundary structures. In particular smaller SMEs could profit from *community clouds* to which different entities contribute with their respective (smaller) infrastructure. Community clouds can either aggregate public clouds or dedicated resource infrastructures.

We may thereby distinguish between private and public community clouds. For example smaller organizations may come together only to pool their resources for building a private community cloud. As opposed to this, resellers such as Zimory may pool cloud resources from different providers and resell them.

Community Clouds as such are still just a vision, though there are already indicators for such development, e.g. through Zimory [12] and RightScale [13]. Community clouds show some overlap with GRIDs technology (see e.g. Reservoir [40]).

Special Purpose Clouds. In particular IaaS clouds originating from data centres have a "general purpose" appeal to them, as their according capabilities can be equally used for a wide scope of use cases and customer types. As opposed to this, PaaS clouds tend to provide functionalities more specialized to specific use cases, which should not be confused with "proprietariness" of the platform: specialization implies providing additional, *use case* specific *methods*, whilst proprietary data implies that *structure* of data and interface are specific to the *provider*.

Specialized functionalities are provided e.g. by the Google App Engine which provides specific capabilities dedicated to distributed document management. Similar to general service provisioning (web based or not), it can be expected that future systems will provide even more specialized capabilities to attract individual user areas, due to competition, customer demand and available expertise.

Special Purpose Clouds are just extensions of "normal" cloud systems to provide additional, dedicated capabilities. The basis of such development is already visible.

3. Cloud Environment Roles

In cloud environments, individual roles can be identified similar to the typical role distribution in Service Oriented Architectures and in particular in (business oriented) Virtual Organisations. As the roles relate strongly to the individual business models it is imperative to have a clear definition of the types of roles involved in order to ensure common understanding.

(Cloud) Providers offer *clouds* to the customer – either via dedicated APIs (PaaS), virtual machines and / or direct access to the resources (IaaS). Note that hosts of cloud enhanced services (SaaS) are typically referred to as *Service Providers*, though there may be ambiguity between the terms Service Provider and Cloud Provider.

(Cloud) Resellers or Aggregators aggregate cloud platforms from *cloud providers* to either provide a larger resource infrastructure to their customers or to provide enhanced features (see II.B). This relates to *community clouds* in so far as the cloud aggregators may expose a single interface to a

merged cloud infrastructure. They will match the economic benefits of global cloud infrastructures with the understanding of local customer needs by providing highly customized, enhanced offerings to local companies (especially SME's) and world-class applications in important European industry sectors. Similar to the software and consulting industry, the creation of European cloud partner eco-systems will provide significant economic opportunities in the application domain – first, by mapping emerging industry requests into innovative solutions and second by utilizing these innovative solutions by European companies in the global marketplace.

(Cloud) Adopters or (Software / Services) Vendors enhance their own services and capabilities by exploiting cloud platforms from *cloud providers* or *cloud resellers*. This enables them to e.g. provide services that scale to dynamic demands – in particular new business entries who cannot estimate the uptake / demand of their services as yet (cf. II.B.1). The cloud enhanced services thus effectively become *software as a service*.

(Cloud) Consumers or Users make *direct* use of the cloud capabilities (cf. below) – as opposed to *cloud resellers* and *cloud adopters*, however, not to improve the services and capabilities they offer, but to make use of the direct results, i.e. either to execute complex computations or to host a flexible data set. Note that this involves in particular larger enterprises which outsource their in-house infrastructure to reduce cost and efforts (see also *hybrid clouds*).

Note that future market developments will most likely enable the user to become provider and consumer at the same time, thus following the "Prosumer" concept, as already introduced by the Service Oriented Architecture concepts [8].

(Cloud) Tool Providers do not actually provide cloud capabilities, but supporting tools such as programming environments, virtual machine management etc.

B. Specific Characteristics / Capabilities of Clouds

Since "clouds" do not refer to a specific technology, but to a general provisioning paradigm with enhanced capabilities, it is mandatory to elaborate on these aspects. There is currently a strong tendency to regard clouds as "just a new name for an old idea", which is mostly due to a confusion between the cloud concepts and the strongly related P/I/SaaS paradigms (see also II.A.2, but also due to the fact that similar aspects have already been addressed without the dedicated term "cloud" associated with it (see also II).

This section specifies the concrete capabilities associated with clouds that are considered *essential* (required in any cloud environment) and *relevant* (ideally supported, but may be restricted to specific use cases). We can thereby distinguish non-functional, economic and technological capabilities addressed, respectively to be addressed by cloud systems.

Non-functional aspects represent *qualities* or *properties* of a system, rather than specific technological requirements. Implicitly, they can be realized in multiple fashions and interpreted in different ways which typically leads to strong compatibility and interoperability issues between individual providers as they pursue their own approaches to realize their respective requirements, which strongly differ between providers. Non-functional aspects are one of the key reasons why "clouds" differ so strongly in their interpretation (see also II.B).

Economic considerations are one of the key reasons to introduce cloud systems in a business environment in the first instance. The particular interest typically lies in the reduction of cost and effort through outsourcing and / or automation of essential resource management. As has been noted in the first section, relevant aspects thereby to consider relate to the cut-off between loss of

control and reduction of effort. With respect to hosting private clouds, the gain through cost reduction has to be carefully balanced with the increased effort to build and run such a system.

Obviously, **technological challenges** implicitly arise from the non-functional and economical aspects, when trying to realize them. As opposed to these aspects, technological challenges typically imply a specific realization – even though there may be no standard approach as yet and deviations may hence arise. In addition to these implicit challenges, one can identify additional technological aspects to be addressed by cloud system, partially as a pre-condition to realize some of the high level features, but partially also as they directly relate to specific characteristics of cloud systems.

1. Non-functional Aspects

The most important non-functional aspects are:

- **Elasticity** is an essential core feature of cloud systems and circumscribes the capability of the underlying infrastructure to adapt to changing, potentially non-functional requirements, for example amount and size of data supported by an application, number of concurrent users etc. One can distinguish between horizontal and vertical scalability, whereby *horizontal scalability* refers to the amount of instances to satisfy e.g. changing amount of requests, and *vertical scalability* refers to the size of the instances themselves and thus implicit to the amount of resources required to maintain the size. Cloud scalability involves both (rapid) up- *and* down-scaling.

Elasticity goes one step further, tough, and does also allow the dynamic integration and extraction of physical resources to the infrastructure. Whilst from the application perspective, this is identical to scaling, from the middleware management perspective this poses additional requirements, in particular regarding reliability. In general, it is assumed that changes in the resource infrastructure are announced first to the middleware manager, but with large scale systems it is vital that such changes can be maintained automatically.

- **Reliability** is essential for all cloud systems – in order to support today's data centre-type applications in a cloud, reliability is considered one of the main features to exploit cloud capabilities. Reliability denotes the capability to ensure constant operation of the system without disruption, i.e. no loss of data, no code reset during execution etc. Reliability is typically achieved through redundant resource utilisation. Interestingly, many of the reliability aspects move from a hardware to a software-based solution. (Redundancy in the file systems vs. RAID controllers, stateless front end servers vs. UPS, etc.).

Notably, there is a strong relationship between availability (see below) and reliability – however, reliability focuses in particular on prevention of loss (of data or execution progress).

- **Quality of Service** support is a relevant capability that is essential in many use cases where specific requirements have to be met by the outsourced services and / or resources. In business cases, basic QoS metrics like response time, throughput etc. must be guaranteed at least, so as to ensure that the quality guarantees of the cloud user are met. *Reliability* is a particular QoS aspect which forms a specific quality requirement.

- **Agility and adaptability** are essential features of cloud systems that strongly relate to the elastic capabilities. It includes on-time reaction to changes in the amount of requests and size of resources, but also adaptation to changes in the environmental conditions that e.g. require different *types* of resources, different *quality* or different *routes,* etc. Implicitly, agility and adaptability require resources (or at least their management) to be autonomic and have to enable them to provide self-* capabilities.

Availability of services and data is an essential capability of cloud systems and was actually one of the core aspects to give rise to clouds in the first instance. It lies in the ability to introduce redundancy for services and data so failures can be masked transparently. Fault tolerance also requires the ability to introduce new redundancy (e.g. previously failed or fresh nodes) in an online manner non-intrusively (without a significant performance penalty).

With increasing concurrent access, availability is particularly achieved through replication of data / services and distributing them across different resources to achieve load-balancing. This can be regarded as the original essence of scalability in cloud systems.

2. Economic Aspects

In order to allow for economic considerations, cloud systems should help in realising the following aspects:

Cost reduction is one of the first concerns to build up a cloud system that can adapt to changing consumer behaviour and reduce cost for infrastructure maintenance and acquisition. *Scalability* and *Pay per Use* are essential aspects of this issue. Notably, setting up a cloud system typically entails additional costs – be it by adapting the business logic to the cloud host specific interfaces or by enhancing the local infrastructure to be "cloud-ready". See also *return of investment* below.

Pay per use. The capability to build up cost according to the actual consumption of resources is a relevant feature of cloud systems. Pay per use strongly relates to quality of service support, where specific requirements to be met by the system and hence to be paid for can be specified. One of the key economic drivers for the current level of interest in cloud computing is the structural change in this domain. By moving from the usual capital upfront investment model to an operational expense, cloud computing promises to enable especially SME's and entrepreneurs to accelerate the development and adoption of innovative solutions.

Improved time to market is essential in particular for small to medium enterprises that want to sell their services quickly and easily with little delays caused by acquiring and setting up the infrastructure, in particular in a scope compatible and competitive with larger industries. Larger enterprises need to be able to publish new capabilities with little overhead to remain competitive. Clouds can support this by providing infrastructures, potentially dedicated to specific use cases that take over essential capabilities to support easy provisioning and thus reduce time to market.

Return of investment (ROI) is essential for all investors and cannot always be guaranteed – in fact some cloud systems currently fail this aspect. Employing a cloud system must ensure that the cost and effort vested into it is outweighed by its benefits to be commercially viable – this may entail direct (e.g. more customers) and indirect (e.g. benefits from advertisements) ROI. Outsourcing resources versus increasing the local infrastructure and employing (private) cloud technologies need therefore to be outweighed and critical cut-off points identified.

Turning CAPEX into OPEX is an implicit, and much argued characteristic of cloud systems, as the actual cost benefit (cf. ROI) is not always clear (see e.g.[9]). Capital expenditure (CAPEX) is required to build up a local infrastructure, but with outsourcing computational resources to cloud systems on demand and scalable, a company will actually spend operational expenditure (OPEX) for provisioning of its capabilities, as it will acquire and use the resources according to operational need.

"Going Green" is relevant not only to reduce additional costs of energy consumption, but also to reduce the carbon footprint. Whilst carbon emission by individual machines can be quite well estimated, this information is actually taken little into consideration when scaling systems up.

Clouds principally allow reducing the consumption of unused resources (down-scaling). In addition, up-scaling should be carefully balanced not only with cost, but also carbon emission issues. Note that beyond software stack aspects, plenty of Green IT issues are subject to development on the hardware level.

3. Technological Aspects

The main technological challenges that can be identified and that are commonly associated with cloud systems are:

Virtualisation is an essential technological characteristic of clouds which hides the technological complexity from the user and enables enhanced flexibility (through aggregation, routing and translation). More concretely, virtualisation supports the following features:

Ease of use: through hiding the complexity of the infrastructure (including management, configuration etc.) virtualisation can make it easier for the user to develop new applications, as well as reduces the overhead for controlling the system.

Infrastructure independency: in principle, virtualisation allows for higher interoperability by making the code platform independent.

Flexibility and Adaptability: by exposing a virtual execution environment, the underlying infrastructure can change more flexible according to different conditions and requirements (assigning more resources, etc.).

Location independence: services can be accessed independent of the physical location of the user and the resource.

Multi-tenancy is a highly essential issue in cloud systems, where the location of code and / or data is principally unknown and the same resource may be assigned to multiple users (potentially at the same time). This affects infrastructure resources as well as data / applications / services that are hosted on shared resources but need to be made available in multiple isolated instances. Classically, all information is maintained in separate databases or tables, yet in more complicated cases information may be concurrently altered, even though maintained for isolated tenants. Multi-tenancy implies a lot of potential issues, ranging from data protection to legislator issues (see section III).

Security, Privacy and Compliance is obviously essential in all systems dealing with potentially sensitive data and code.

Data Management is an essential aspect in particular for storage clouds, where data is flexibly distributed across multiple resources. Implicitly, data consistency needs to be maintained over a wide distribution of *replicated* data sources. At the same time, the system always needs to be aware of the data location (when replicating across data centres) taking latencies and particularly work-load into consideration. As size of data may change at any time, data management addresses both horizontal and vertical aspects of scalability. Another crucial aspect of data management is the provided consistency guarantees (eventual vs. strong consistency, transactional isolation vs. no isolation, atomic operations over individual data items vs. multiple data times etc.).

APIs and / or Programming Enhancements are essential to exploit the cloud features: common programming models require that the developer takes care of the scalability and autonomic capabilities him- / herself, whilst a cloud environment provides the features in a fashion that allows the user to leave such management to the system.

Metering of any kind of resource and service consumption is essential in order to offer elastic pricing, charging and billing. It is therefore a pre-condition for the elasticity of clouds.

Tools are generally necessary to support development, adaptation and usage of cloud services.

C. Related Areas

It has been noted, that the cloud concept is strongly related to many other initiatives in the area of the "Future Internet", such as Software as a Service and Service Oriented Architecture. New concepts and terminologies often bear the risk that they seemingly supersede preceding work and thus require a "fresh start", where plenty of the existing results are lost and essential work is repeated unnecessarily. In order to reduce this risk, this section provides a quick summary of the main related areas and their potential impact on further cloud developments.

1. Internet of Services

Service based application provisioning is part of the Future Internet as such and therefore a similar statement applies to cloud and Internet of Services as to cloud and Future Internet. Whilst the cloud concept foresees essential support for service provisioning (making them scalable, providing a simple API for development etc.), its main focus does not primarily rest on service provisioning. As detailed in section II.A.1 cloud systems are particularly concerned with providing an infrastructure on which any type of service can be executed with enhanced features.

Clouds can therefore be regarded as an enabler for enhanced features of large scale service provisioning. Much research was vested into providing base capabilities for service provisioning – accordingly, capabilities that overlap with cloud system features can be easily exploited for cloud infrastructures.

2. Internet of Things

It is up to debate whether the Internet of Things is related to cloud systems at all: whilst the internet of things will certainly have to deal with issues related to elasticity, reliability and data management etc., there is an implicit assumption that resources in cloud computing are of a type that can host and / or process data – in particular storage and processors that can form a computational unit (a virtual processing platform).

However, specialised clouds may e.g. integrate dedicated sensors to provide enhanced capabilities and the issues related to reliability of data streams etc. are principally independent of the type of data source. Though sensors as yet do not pose essential scalability issues, metering of resources will already require some degree of sensor information integration into the cloud.

Clouds may furthermore offer vital support to the internet of things, in order to deal with a flexible amount of data originating from the diversity of sensors and "things". Similarly, cloud concepts for scalability and elasticity may be of interest for the internet of things in order to better cope with dynamically scaling data streams.

Overall, the Internet of Things may profit from cloud systems, but there is no direct relationship between the two areas. There are however contact points that should not be disregarded. Data management and interfaces between sensors and cloud systems therefore show commonalities.

3. The Grid

There is an on-going confusion about the relationship between Grids and Clouds [17], sometimes seeing Grids as "on top of" Clouds, vice versa or even identical. More surprising, even elaborate comparisons (such as [18][19][20]) still have different views on what "the Grid" is in the first

instance, thus making the comparison cumbersome. Indeed most ambiguities can be quickly resolved if the underlying concept of Grids is examined first: just like Clouds, Grid is primarily a concept rather than a technology thus leading to many potential misunderstandings between individual communities.

With respect to research being carried out in the Grid over the last years, it is therefore recommendable to distinguish (at least) between (1) "Resource Grids", including in particular Grid Computing, and (2) "eBusiness Grids" which centres mainly on distributed Virtual Organizations and is closer related to Service Oriented Architectures (see below). Note that there may be combination between the two, e.g. when capabilities of the eBusiness Grids are applied for commercial resource provisioning, but this has little impact on the assessment below.

Resource Grids try to make resource - such as computational devices and storage - locally available in a fashion that is transparent to the user. The main focus thereby lies on *availability* rather than scalability, in particular rather than *dynamic* scalability. In this context we may have to distinguish between HPC Grids, such as EGEE, which select and provide access to (single) HPC resources, as opposed to distributed computing Grids (cf. Service Oriented Architecture below) which also includes P2P like scalability - in other words, the more resources are available, the more code instances are deployed and executed. Replication capabilities may be applied to ensure *reliability*, though this is not an intrinsic capability of in particular computational Grids. Even though such Grid middleware(s) offers manageability interfaces, it typically acts on a layer on top of the actual resources and thus does rarely virtualise the hardware, but the computing resource as a whole (i.e. not on the IaaS level).

Overall, Resource Grids do address similar issues to Cloud Systems, yet typically on a different layer with a different focus - as such, e.g. Grids do generally not cater for horizontal and vertical elasticity. What is more important though is the strong conceptual overlap between the issues addressed by Grid and Clouds which allows re-usage of concepts and architectures, but also of parts of technology (see also SOA below).

Specific shared concepts:

- Virtualisation of computation resources, respectively of hardware
- Scalability of amount of resources versus of hardware, code and data
- Reliability through replication and check-pointing
- Interoperability
- Security and Authentication

eBusiness Grids share the essential goals with Service Oriented Architecture, though the specific focus rests on integration of existing services so as to build up new functionalities, and to enhance these services with business specific capabilities. The eBusiness (or here "Virtual Organization") approach derives in particular from the distributed computing aspect of Grids, where parts of the overall logic are located in different sites. The typical Grid middleware thereby focus mostly on achieving reliability in the *overall* execution through on-the-fly replacement and (re)integration.

But eBusiness Grids also explore the specific requirements for commercial employment of service consumption and provisioning - even though this is generally considered an aspect more related to Service Oriented Architectures than to Grids.

Again, eBusiness Grids and Cloud Systems share common concepts and thus basic technological approaches. In particular with the underlying SOA based structure, capabilities may be exposed and integrated as stand-alone services, thus supporting the re-use aspect.

Specific shared concepts:

- Pay-per-use / Payment models
- Quality of Service
- Metering
- Availability through self-management

It is worth noting that the comparison here is with deployed Grids. The original Grids concept had a vision of elasticity, virtualization and accessibility [48] [49] not unlike that claimed for the Clouds vision.

4. Service Oriented Architectures

There is a strong relationship between the "Grid" and Service Oriented Architectures, often leading to confusions where the two terms either are used indistinguishably, or the one as building on top of the other. This arises mostly from the fact that both concepts tend to cover a comparatively wide scope of issues, i.e. the term being used a bit ambiguously.

Service Oriented Architecture however typically focuses predominantly on ways of developing, publishing and integrating application logic and / or resources as services. Aspects related to enhancing the provisioning model, e.g. through secure communication channels, QoS guaranteed maintenance of services etc. come in this definition secondary. Again it must be stressed though that the aspects of eBusiness Grids and SOA are used almost interchangeably - in particular since the advent of Web Service technologies such as the .NET Framework and Globus Toolkit 4, where GT4 is typically regarded as Grid related and .NET as a Web Service / SOA framework (even though they share the same main capabilities).

Though providing cloud hosted applications as a service is an implicit aspect of Cloud SaaS provisioning, the cloud concept is principally technology agnostic, but it is generally recommended to build on service-oriented principles. However, in particular with the resource virtualization aspect of cloud systems, most technological aspects will have to be addressed at a lower level than the service layer.

Service Oriented Architectures are therefore of primary interest for a) the type of applications and services the user can build for and host on the cloud system and b) for providing additional high-level services and capabilities with which to enhance the base cloud capabilities.

III. State of the Art & Analysis

As has been noted in the preceding section, cloud systems are not a new technology yet to be developed – instead plenty of existing technologies branded the name to demark specific capabilities and concepts. Accordingly, relevant progress has already been made on both the commercial and the academic side of cloud systems. What is more, with the relationship of clouds to other research areas, as elaborated in section II.C, substantial results are available that will directly impact on future development and research in the area of cloud technologies.

This section will examine the current state of play in the area of cloud systems as a foundation for future research and development. It should be noted in this context that, so far, primarily few commercial companies have invested into specific progress in the area of global cloud technologies, as the according infrastructure is typically too costly for small to medium players and of less essential relevance for their business models. In particular infrastructure providers have vested substantial efforts into (autonomous) maintenance of their resources, thus laying the foundation for cloud systems. As opposed to this, research in related areas and academic research based on other funding principles and interests therefore contributed to cloud technologies in particularly indirectly so far.

The overview over state of the art therefore distinguishes between commercial and academic / research focused efforts.

A. Current Commercial Efforts

The most well-known commercial cloud providers, implementing at least significant parts of the concept described in Part A, are Amazon, Google and Force.com – not alone for the reason that they mainly coined the term "cloud" for the respective set of functionalities and capabilities offered, even though their functional scope already differs substantially (see also section I).

It has to be noted that commercial efforts are driven by other motivations than publically sponsored research initiatives and act on different timescales. Industrial efforts are customer and result driven and focus on sustainable return of investment rather than technological convergence per se. (The significant upfront investments are in opposition to "quick" ROI models).

This section does not try to detail all the commercial models currently available (please refer to e.g. [10] for a more exhaustive overview), but to capture the most relevant technological advances made in these areas with respect to cloud systems. In other words, it tries to summarise the main support that both new providers and customers (including aggregators) can acquire through commercial tools.

The following tables provide an overview over the main features that uptakers can expect from current commercial tools to the authors best knowledge, thereby following the same structure as introduced in section II.B, regarding the main capabilities of cloud systems. The tables read as follows: the main capabilities (row) are met / failed by commercial products in the way designated in the columns, whereas a tick ☑ implies that the respective feature / capability is provided through some existing tools (i.e. is no unsolved issues in the according domain) and an empty box ☐ denotes aspects that are considered relevant in the respective context but are not well supported. Columns 4-7 denote the specific support a new provider (columns 4-6) can expect through commercial tools, respectively the specific capabilities a cloud user (column 7) can expect from commercial cloud providers.

1. Non-Functional Aspects Overview

	General	Examples	(IaaS)	(PaaS)	(SaaS)	(Users)
Elasticity	☑ horizontal scale-out ☐ vertical scalability ☐ efficient scale-down	horizontal: Amazon EC2 ; Amazon S3; Google Docs; eBay, MS Azure vertical: Xen; Amazon S3 (to a degree)	☑ horizontal scale ☐ vertical scale ☐ scale-down	☑ horizontal scale	☑ horizontal scale	☑ scalability ☐ potentially too high resource consumption
Reliability	☑ reliable data storage - no code execution	Xen Server Virtualisation, VMWare	☑ reliable data storage ☐ no code execution	☑ reliable app execution	☑ reliable data storage ☐ no code execution	☑ data replication
Quality of Service	☑ resource level QoS solved ☐ little usage in clouds ☐ no higher level representation	Cisco, Amazon S3, Amazon EC2	☑ resource level QoS ☐ no abstraction	☐ no SLA	☐ hardly any SLA	☑ basic quality guarantees
Agility and adaptability	☑ see elasticity ☐ little adaptability to use cases ☐ little adaptability to technology	RightScale, FlexNet	☑ adapt to resource (virtualisation) ☐ only on image level	☑ elasticity ☐ static APIs	☑ elasticity ☐ depends fully on service' capabilities	☐ has to adapt code to system not vice versa
Availability	☑ high availability ☐ basically only through replication ☐ requires large infrastructure	MS Azure, Amazon S3	☑ high data availability ☐ little resource availability	☑ high data availability ☑ fair applet availability	☑ high data availability Note: service availability depends on complexity	☑ data availability ☐ service availability ☐ resource availability

Table 1: Non-functional aspects addressed by current commercial efforts (☑ supported; ☐ deficiency)

2. Economic Aspects Overview

	General	Examples	(IaaS)	(PaaS)	(SaaS)	(Users)
Cost reduction	☑ simplified service provisioning ☑ simplified resource management ☐ proprietary structures ☐ no general recommendations (cf. "improved time to market")	Google Apps Engine (through scaling)	☑ resource management ☐ no general rules	☑ resource mgmt ☑ scale management ☐ recommendations	☑ resource & scaling management ☐ no general policies	☑ outsourcing ☑ reduced mgmt overhead ☑ scalability ☐ change vs. gain ☐ too high resource consumption
			all providers have the full costs of providing and maintaining the resources - cost reduction is mostly on user's side.			
Pay per use	☑ static billing ☑ dynamicity e.g. in DSL ☐ use case specific ☐ not related to resource availability	PayPal, HP PPU	☑ basic billing support ☐ little resource specific support ☐ no relationship to QoS management	☑ basic billing support ☐ little service specific support	☑ basic billing support ☐ little service specific support	☑ automatic billing ☐ little negotiation support ☐ little QoS related support
Improved time to market	☑ simplified service provisioning ☑ simplified resource management ☐ proprietary structures	Animoto	n/a	n/a	n/a	☑ simplified resource & service lifecycle ☑ simple (use case specific) APIs ☐ use case specific ☐ vendor lock-in
			applies only to aggregators,resellers or consumers			
Return of investment (ROI)	☑ outsourcing & work offloading ☐ difficult to assess ☐ no general guidelines		☐ no general recommendations	☐ no general recommendations	☐ no general recommendations	☑ outsourcing & work offloading ☐ general guidelines
			applies mostly to cloud uptakers			
Turning CAPEX into OPEX	*General issue*		No dedicated tool support			
"Going Green"	☑ addressed by data centres ☑ EC code of conduct [21] ☐ little support "in the cloud"	EfficientServers	☑ measurement mechanisms ☑ EC code of conduct ☑ greener hardware (e.g. Intel Atom) ☐ needs to be implemented manually	☑ EC code of conduct ☐ needs to be implemented manually	☑ EC code of conduct ☐ needs to be implemented manually	☑ outsourcing ☑ dynamic scalability ☐ effectively manually

Table 2: Economical aspects addressed by current commercial efforts (☑ supported; ☐ deficiency)

3. Technological Aspects Overview

	General	Examples	(IaaS)	(PaaS)	(SaaS)	(Users)
Virtualisation	☑ some virtualisation in all clouds ☑ numerous technologies ☑ location independence ☐ difficult to use ☐ no interoperability	Xen, Virtual PC, VMWare, Virtual Box, MS HyperV	☑ machine virtualisation ☑ routing, security … ☑ leave images to customer ☐ only images	☑ easier resource maintenance ☑ routing ☐ difficult to use	☑ easier resource maintenance ☑ routing ☐ difficult to use	☑ simple access ☐ no interoperability
Multi-tenancy	☑ general data management support ☐ little multi-purpose solutions	MS SQL [27]	☑ image separation ☐ VM support little cross resource multi-tenancy issues	☑ general data management support ☑ engine re-usage ☐ mostly manual	☑ data mgmt. ☑ instantiation support ☐ manual	☑ higher availability ☐ data consistency manual (see data management)
Security and Compliance	☑ encryption ☑ identification, authentication & authorization ☑ data rights management ☐ legislative regulation ☐ constant changes ☐ compliance with specific security requirements	almost all	☑ encryption, authentication etc. ☑ virtual machine separation ☐ only valid for access portals	☑ encryption, authentication etc. Note: manual configuration but only per engine	☑ encryption, authentication etc. ☐ manual configuration per service	☑ easily available ☑ mostly catered for by provider ☐ legislative regulations not available / not observed
Data Management	☑ many basic issues addressed ☑ distributed data management ☑ versioning ☑ conversion ☐ always new challenges ☐ little interoperability ☐ consistency, scalability, growth	Mesh, Amazon Dynamo, WebSphere	☑ general data management support ☐ no specific data management across virtual machines ☐ efficiency	☑ general data management support ☐ consistency management ☐ concurrency ☐ efficiency	☑ general data management support ☐ consistency management ☐ concurrency ☐ efficiency	☑ data available anywhere ☐ consistency mostly manual ☐ little interoperability - speed vs. size
APIs and / or Programming Enhancements	☑ use case specific "simple" APIs ☑ generic programming models ☑ full application development for clouds ☐ complexity ☐ control	MS Azure, Google App Engine, Hadoop	n/a	☑ use case specific APIs (engines) ☐ complexity ☐ control	☑ generic programming models ☐ complexity ☐ control	☑ different programming models ☐ complexity mostly with the developer ☐ little in-depth control

Table 3: Technological aspects addressed by current commercial efforts (☑ supported; ☐ deficiency)

4. Assessment

Overall, public clouds of the types introduced in section II.A.1 are commercially available - a more exhaustive comparison of existing providers and their features at the time of writing is available through Webhosting Unleashed [34] and Infoworld.com [35]. Current cloud systems still suffer a lot of drawbacks and do not overall offer the infrastructure expected to be required in the near future - this relates in particular to the typical topics in the IT area, i.e. Data Management, Privacy & Security, Virtualisation and Resource Control (see section III.C.1).

At the same time, existing infrastructures will be difficult to change to new technologies and / or conceptual approaches, making long-term interoperability and standardisation efforts difficult – whereby standardization typically follows interoperability efforts in the commercial domain. But this also poses problems on modelling the policies and dynamic aspects of resource management (see e.g. [22]). Implicitly, non-technical aspects, such as restrictions due to Legislation & Policies, but also Economical Concerns related to whether the move to a cloud infrastructure is economically feasible are of major concern for commercial providers (see section III.C.2).

A currently recurring issue in the context of commercial cloud provisioning consists in "vendor lock-in": As most commercial tools were developed independently from one another with a particular focus on solving the respective company's customers' problems first, there is little (technical) convergence between the available products. This is also due to the typical development cycle of clouds which typically start as in-house, internal solutions (private clouds) which are then extended to provide (a subset of) capabilities to potential customers (public clouds). Issues related to Federation & Interoperability are hence a specific issue for commercial cloud systems (see section III.C.1 "Federation & Interoperability").

An attempt to set up an open cloud forum to counteract the effect of lock-ins basically failed when in particular larger vendors' strongly expressed their desire to perpetuate the lock-in for competition reasons, even though multiple companies still signed the Open Cloud Manifesto [23]. Given the scope of cloud types (cf. section II.A.1), interoperability is however not an issues easily solved by agreeing on common interfaces, as it impacts on different technologies (such as interfaces for SaaS, APIs for PaaS and images for IaaS) – hence it remains dubious whether approaches such as standardization or the Open Cloud Manifesto can actually solve the problem of vendor lock-in [24].

In general, essential support for specific use cases with minor requirements towards the cloud infrastructure can already be provided through commercial tools. However, the available tools and systems are typically restricted to specific use cases which implicitly form the capability support of these tools. It is to be expected that future use cases (see also IV.B.2) will put forward higher demands towards the scope of these capabilities which is not currently met.

B. Current Research

So far, only few cloud dedicated research projects in the widest sense have been initiated – most prominent amongst them probably OpenNebula and Reservoir. However, many projects have initiated a dedicated cloud related research track investigating into how to move existing capabilities onto and into the cloud. What is more, countless projects have addressed similar concepts in related areas (see II.C) exhaustively and have provided relevant results that need to be taken up in order to exploit relevant intellectual results, as well as to ensure that no effort is unnecessarily repeated, thus reducing the chance for impact and uptake. It is notable in this context, that uptake of research results is generally slow, in particular in comparison to commercial results.

Just like with the preceding section on current commercial efforts, the following tables provide an overview over the current status of research efforts with respect to the capabilities assigned to cloud systems (section II.B). The tables follow the same structure as in the preceding section, i.e. they list the main capabilities per characteristic supported, respectively failed through general research efforts at the moment.

1. Non-Functional Aspects Overview

	General	Examples	(IaaS)	(PaaS)	(SaaS)	(Users)
Elasticity	☑ horizontal scale-out ☑ limited vertical scale-out ☐ efficient scale-down	XenBEE	☑ horizontal scale ☑ vertical scale(offline mode) ☐ efficient scale-down	☑ horizontal scale ☐ vertical scale ☐ scale-down	☑ horizontal scale ☐ no vertical scale ☐ efficient scale-down	☑ scalability ☑ limited vertical scalability - resource consumption
Reliability	☑ reliable data storage ☑ early failure warning ☑ code execution replication ☐ no actual reliable code execution yet	PHASTGrid, GWES	☑ reliable storage ☑ early warning ☑ code replication and check-pointing ☐ code execution support	☑ reliable app execution ☑ early resource failure detection ☑ replication	☑ reliable app execution ☑ early resource failure detection ☑ replication	☑ data reliability ☐ limited code reliability
Quality of Service	☑ QoS definition and enforcement across all tiers ☑ limited negotiation, optimisation and abstraction ☐ effective scheduling ☐ QoS based self-*	TrustCoM, BREIN, SLA@SOI	☑ QoS management on resource level ☐ effective scheduling ☐ adaptation according to QoS	☑ QoS on service and resource level ☑ limited negotiation ☐ effective scheduling ☐ adaptation	☑ QoS on service and resource level ☑ limited negotiation ☐ effective scheduling ☐ adaptation	☑ QoS monitoring and enforcement ☐ only limited negotiation and abstraction
Agility and adaptability	☑ see elasticity ☑ limited (self)awareness ☑ use case specific reasoning ☐ limited to use case ☐ limited to specific technology	TIMaCS, GWES, VieSLAF	☑ adapt to resource (virtualisation) ☑ some resource self-adaptation ☐ use case specific	☑ elasticity ☑ some self-* ☑ some reasoning ☐ limited to specific technology	☑ elasticity ☑ some self-awareness and ☐adaptation ☐ limited to specific technology	☑ some intelligent behaviour ☐ has to adapt code to system not vice versa
Availability	☑ availability of all types of resources and services ☑ routing, virtualisation, connectivity ☑ complex scheduling with wait time ☐ on-demand / on-the-fly scheduling ☐ compensating insufficient resources	OpenNebula, EGEE, PHASTGrid	☑ general availability through virtualization ☑ complex scheduling ☐ compensating insufficient resources	☑ general availability ☑ routing ☑ complex scheduling ☐ on-demand scheduling	☑ general availability ☑ routing ☑ complex scheduling ☐ on-demand scheduling	☑ general availability ☐ compensating insufficient resources

Table 4: Non-functional aspects addressed by current research efforts (☑ supported; ☐ deficiency)

	General	Examples	(IaaS)	(PaaS)	(SaaS)	(Users)
Cost reduction	☑ more efficient resource usage ☑ resource and service provisioning / usage ☑ policy systems support outsourcing decision ☐ no general economical recommendations ☐ optimisation		☑ efficient resource usage ☑ policy based self-* ☐ no general recommendations ☐ optimisation	☑ resource management ☑ scaling management ☑ policy based self-* ☐ no general recommendations ☐ optimisation	☑ resource management ☑ scaling management ☑ policy based self-* ☐ no general recommendations ☐ optimisation	☑ outsourcing ☑ reduced mgmt overhead ☑ scalability ☐ effort vs. gain ☐ potentially too high resource consumption
Pay per use	☑ SLA / QoS based metering ☑ access & consumption based billing	SLA@SOI, TrustCoM, Gria, Nagios, Ganglia	☑ SLA related support ☐ only on resource level (not generally in image) (see SLA)	☑ SLA related support (see SLA)	☑ SLA related support (see SLA)	☑ SLA related support ☐ no abstraction / aggregation of cost (see SLA)
Improved time to market	☐ highly use case dependent Note: time to market is generally improved thanks to scalability and availability		n/a	n/a	n/a	☑ simplified resource & service lifecycle ☑ simple (use case specific) APIs - use case specific
Return of investment (ROI)	☑ policy systems can regulate the decision ☐ no general policies /recommendations		☐ general recommendations	☐ general recommendations	☐ general recommendations	☑ outsourcing & work offloading ☑ policy based support ☐ general guidelines
Turning CAPEX into OPEX	*general issue*		*No dedicated tool support*			
"Going Green"	☑ increased interest ☑ policy based rules ☑ manageable resource ☐ no "green" manageability ☐ no "green" scheduling ☐ little policies / recommendations		☑ measurement mechanisms ☑ greener hardware ☑ some hardware level mechanisms ☐ mostly manual	☐ mostly manual	☐ mostly manual	☑ outsourcing ☑ dynamic scalability ☐ mostly manual

TABLE 5: ECONOMICAL ASPECTS ADDRESSED BY CURRENT RESEARCH EFFORTS (☑ SUPPORTED; ☐ DEFICIENCY)

	General	Examples	(IaaS)	(PaaS)	(SaaS)	(Users)
Virtualisation	☑ numerous virtualisation technol. ☑ all tiers ☑ commercial-like open source products ☐ limited control ☐ difficult to use and manage ☐ proprietary structures	IRMOS, XenBEE	☑ machine virtualization ☑ routing, sec. etc. ☑ leave images to customer ☐ restricted to images ☐ proprietary structs.	☑ service virtualization ☑ routing, security etc. ☐ proprietary structures ☐ difficult to use and manage	☑ service virtualization ☑ routing, security etc. ☐ proprietary structures ☐ difficult to use and manage	☑ simpler access ☑ hidden complexity ☐ limited interoperability
Multi-tenancy	☑ virtual machine like separation ☑ data handling with various protection modes ☐ data locking may occur		☑ image check-pointing etc. ☐ little cross resource multi-tenancy support	☑ engine re-usage ☑ data handling with various protection modes ☐ data locking	☑ policy based instantiation support ☑ data handling - data locking	☑ higher availability ☑ some consistency management
Security and Compliance	☑ base security issues covered ☑ federated identities ☐ new security holes ☐ legislation related aspects	MS Geneva, BREIN, RESERVOIR	☑ base security covered ☑ VM separation ☐ valid for portals, no general ctrl in VMs	☑ base security ☐ manual configuration but only per engine	☑ base security ☑ semi-automatic configuration	☑ easily available ☑ mostly catered for by provider ☐ legislative regulation issues
Data Management	☑ base issues addressed ☑ distributed data management ☑ versioning, visualisation etc. ☐ little consistency / conflict resolution ☐ efficient data size management ☐ little efficient segmentation and distribution	OGSA-DAI, iRods, SRB, LarkC	☑ general data management support ☐ no specific data management across virtual machines ☐ efficiency	☑ general data management support ☑ some consistency support (use case specific) ☐ concurrency ☐ efficiency	☑ general data management support ☑ some consistency support (use case specific) ☐ consistency mgmt ☐ concurrency ☐ efficiency	☑ data available anywhere ☑ versioning etc. ☐ consistency mostly manual ☐ little interoperability ☐ speed vs. size
APIs and / or Programming Enhancements	☑ distributed programming language ☐ HPC focus ☐ little ease-of-use ☐ little flexibility	MPI, PGAS (UPC, CAF, Chapel, X10), ParallelC#	n/a	☑ use case specific APIs (engines) ☐ complexity ☐ control	☑ some self-distributing programming models ☑ some resource control ☐ complexity	☑ different programming models ☐ complexity mostly with the developer ☐ little in-depth control

TABLE 6: TECHNOLOGICAL ASPECTS ADDRESSED BY CURRENT RESEARCH EFFORTS (☑ SUPPORTED; ☐ DEFICIENCY)

4. Assessment

Research and the open source development community typically centre on individual capabilities rather than integrated systems and holistic middleware - accordingly, it is not surprising that many of the available results consist in tools with dedicated capabilities. These tools are sometimes aligned with other systems and tools, if part of a larger research project. There are only a few large-scope, more generic frameworks for cloud systems, such as OpenNebula which concentrates on a virtualization layer for IaaS though.

Notably a complete infrastructure system may not even be in the interest of the (research) community or of open source uptakers, as they tend towards proprietary data structures and interfaces in order to compensate for gaps in specifications and existing tools. In other words, it may be more sensible to consider development of whole infrastructures an integration task over existing tools, rather than a standalone RTD issue.

Most research results adhere to SOA paradigms and try to maintain standard interfaces, mostly basing on Web Service specifications. Thus research results show much higher interoperability than commercial results, which is reflected in the vendor lock-in problem.

However, the stability of research results is still questionable, in particular if used in a wider and more commercially oriented environment. Whilst individual capabilities are supported quite well, it is difficult for a potential user to employ these capabilities in his / her respective environment and adhering to the according requirements. This holds particularly true if capabilities should be combined, i.e. if multiple tools are to be employed in order to meet the requirements. Since most tools have been developed in a historical setting oriented to other use-cases and since cloud systems offer a broad principle scope, most techniques will simply not fit in the respective field. For example, most virtualization technologies aim at the resource level, but not at the hardware level, so that re-usage for cloud purposes is impossible.

Overall, research has made considerable conceptual advances covering most of the fundaments of cloud systems, yet the according technologies and development are mostly lagging behind (see details below). One can thus say, that in all technical areas (section III.C.1), a technological basis has been realized but that still considerable open issues remain in particular due to the additional requirements put forward by cloud applications - these relate specifically to the high degree of scalability as an intrinsic capability of cloud systems. What is more, however, economical issues related to legislative regulations, policies (section III.C.2 “Legislation, Government & Policies”) and how to ensure return of investment, calculation of maximum scalability, quality recommendations etc. (section III.C.2 “Economic Concerns”) have hardly been addressed in research, as they are primarily of commercial concern.

C. Gaps & Open Areas

There is no full scale middleware existent which commonly addresses all cloud capabilities. What is more, not all capabilities can as yet be fulfilled to the necessary extend, even though an essential basis has been provided from both commercial and academic side. The current set of capabilities fulfils the requirements to realise simple cloud systems (as was to be expected given their availability on the market). The particular issue of interest thereby is in how far the available support fulfils the expectations towards cloud systems in their various appearances and use cases (cf. section II).

The main gaps that can be identified relate to the following aspects:

1. Technical Gaps

Manageability and Self-*

Cloud systems focus on intelligent resource management so as to ensure availability of services through their replication and distribution. In principle, this ensures that the amount of resources consumed per service / application reflects the degree of consumption, such as access through users, size of data etc. Whilst most cloud system allow for main features related to elasticity and availability (see Table 1 and Table 4 above), the management features are nowhere near optimal resource usage – issues not only relevant for cost reduction, but also for meeting the green agenda and for ensuring availability when resources are limited.

Management features are mostly use-case specific at the moment and generally better at managing scale-up (e.g. when bandwidth usage exceeds a threshold) than at scale-down (mostly because the duration of inactivity is unpredictable). There is little general support in particular for new providers with respect to how to manage resources, when to scale, how to meet the requirements of the user regarding quality of service etc.

This also involves self-detection of failures, of resource-shortage, but also of free load etc. and taking according actions – in particular in hybrid environments where management has to act across different resource infrastructures and can generally not be centralized. A major criterion thereby consists in improving the performance of management.

Obviously, interoperability plays a major role in distributed management across resource environments, but also the capability to adapt to changes in the environment – this does not only apply to customer requirements (see above), but also to technological restrictions, such as related to relevant libraries (IaaS & SaaS) or engines (PaaS). Adaptability and interoperability are thereby strongly linked to each other.

Management and manageability plays a major role in many of the core cloud characteristics (see e.g. “Elasticity”, “Quality of Service”, “Adaptability” etc. (Table 1, Table 4), and “Cost Reduction”, “Going Green” etc. (Table 2, Table 5), but also implicitly “Data Management” and “Programming Models” (Table 3 and Table 6).

Main issues: efficiency; interoperability; compensating insufficient resources; boundary criteria.

Data Management

The amount of data available on the web, as well as the throughput produced by applications, sensors etc. increases faster than storage and in particular bandwidth does. There is a strong tendency to host more and more public data sets in cloud infrastructures so that improved means of managing and structuring the size of data will be necessary to deal with future requirements. Hence in particular storage clouds should be able to cater for such means in order to maintain availability of data and thus address quality requirements etc.

Not only data size poses a problem for cloud systems, but more importantly consistency maintenance (see section III on “Data Management”), in particular when scaling up. As data may be shared between tenants partially or completely, i.e. either because the whole database is replicated or indeed a subset is subject to concurrent access (such as state information), maintaining consistency over a potentially unlimited number of data instances becomes more and more important and difficult (cf. section III on “Multi-tenancy”). One of the main research gaps and efforts in the area is how to provide truly transactional guarantees for software stacks (e.g. multi-tier

architectures as SAP NetWeaver, Microsoft .NET or IBM WebSphere) that provides large scalability (100s of nodes) without resorting to data partitioning or relaxed consistency (such as eventual consistency). Clearly ACID 2-phase commit transactions will not work (timing) and compensating transactions will be very complex. Worse, the use of caching on distributed database systems means we have to validate cache coherency.

At the moment, segmentation and distribution of data occurs more or less uncontrolled, thus not only leading to efficiency issues and (re)integration problems (see section III on "Data Management"), but also potentially to clashes with legislation (cf. below). In order to be *able* to compensate this, further control capabilities over distribution in the infrastructure are required that allow for context analysis (e.g. location) and QoS fulfilment (e.g. connectivity) - an aspect that is hardly addressed by commercial and / or research approaches so far (see section III on "Elasticity").

As most data in the web is unstructured and heterogeneous due to various data sources, sensible segmentation and usage information requires new forms of annotation. What is more, consistency maintenance strategies may vary between data formats, which can only be compensated by maintaining meta-information about usage and structure. But also with the proprietary structures of individual cloud systems, moving data (and / or services) between these infrastructures is sometimes complicated, necessitating new standards to improve and guarantee long term interoperability (see section III.A.4). Work on the "eXternal Data Representation" (XDR) standard for loosely coupled systems will play an important role in this context.

Cloud resources are potentially shared between multiple tenants – this does not only apply to storage (and CPUs, see below), but potentially also to data (where e.g. a database is shared between multiple users) so that not only changes can occur at different locations, but also in a concurrent fashion. This necessitates improved means to deal with multi-tenancy in distributed data systems.

Classical data management systems break down with large numbers of nodes – even if clustered in a cloud. The latency of accessing disks means that classical transaction handling (two-phase commit) is unlikely to be sustainable if it is necessary to maintain an integral part of the system global state. Efficiency efforts (such as caching) compound the problem needing cache coherency across a very large number of nodes. As current clouds typically use either centralized Storage Area Networks (e.g. Amazon EBS), unshared local disk (e.g. Amazon AMI) or cluster file-systems (e.g. GFS; but for files, not entire disk images), commodity storage (such as desktop PCs) can currently not be easily integrated into cloud storage, even though Live Mesh already allows for synchronization of local storage in / with the cloud.

In order to address these issues, the actual usage behaviour with respect to file and data access in cloud systems need to be assessed more carefully. There are only few of these studies currently available (e.g. [28]), but the according information would help identifying the typical distribution, access, consistency etc. requirements of the individual use cases.

See Table 3 and Table 6, "Data Management" for an overview.

Main issues: data size; interoperability; control; distribution; consistency & multi-tenancy.

Privacy & Security

Strongly related to the issues concerning legislation and data distribution is the concern of data protection and other potential security holes arising from the fact that the resources are shared between multiple tenants and the location of the resources being potentially unknown. In particular sensitive data or protected applications are critical for outsourcing issues. In some use cases, the

information that a certain industry is using the infrastructure at all is enough information for industrial espionage.

Whilst essential security aspects are addressed by most tools, additional issues apply through the specifics of cloud systems, in particular related to the replication and distribution of data in potentially worldwide resource infrastructures. Whilst the data should be protected in a form that addresses legislative issues with respect to data location, it should at the same still be manageable by the system.

In addition, the many usages of cloud systems and the variety of cloud types imply different security models and requirements by the user. As such, classical authentication models may be insufficient to distinguish between the Aggregators / Vendors and the actual User, in particular in IaaS cloud systems, where the computational image may host services that are made accessible to users.

In particular in cases of aggregation and resale of cloud systems, the mix of security mechanisms may not only lead to problems of compatibility, but may also lead to the user distrusting the model due to lack of insight.

All in all, new security governance models & processes are required that cater for the specific issues arising from the cloud model (see also [54]).

See in particular Table 3 and Table 6, for issues concerning “Security and Compliance”.

Main issues: multi-tenancy, trust, data-encryption, legislation compliance.

Federation & Interoperability

One of the most pressing issues with respect to cloud computing is the current difference between the individual vendor approaches, and the implicit lack of interoperability. Whilst a distributed data environment (IaaS) cannot be easily moved to any platform provider (PaaS) and may even cause problems to be used by a specific service (SaaS), it is also almost impossible to move a service / image / environment between providers on the same level (e.g. from Force.com to Amazon).

This issue is mostly caused by the proprietary data structures employed by each provider individually. History of web service standardisation has shown that specifications may easily diverge rather than converge if too many parallel standardisation strands are pursued. Therefore, current standardisation approaches in the web service domain may prove insufficient to deal with the complexity of the problem, as it tends to be slow and diverging between multiple instances of standardization bodies. Also, interoperability is typically driven stronger by de facto standards, than by other de jure standardization efforts.

In particular cloud computing with the strong industrial drivers and the initial uptake already in place has a strong tendency to impel de-facto standards (see also vendor lock in). Traditionally, US – with an emphasis on software innovation - favour a voluntary, market driven approach to standardisation. Europe, with a strong track record in telecom standardisation, seems to favour an upfront approach – albeit mostly in hardware related fields.

While innovations between domains usually benefit from an early focus on interoperability, the quest for disruptive innovations within domains benefits from a lower focus on interoperability requirements in this early phase. Too early focus on interoperability and standardization issues may therefore be disruptive as e.g. long-term requirements and structures cannot be assessed to their full extend today, and a bad specification may hinder interoperable development accordingly. A particular focus must hence rest on atomic, minimal, composable and adaptable standards.

While nobody is questioning the usefulness and benefit of interoperability, it should also be noted that with respect to the European research agenda, careful consideration is necessary in which fields and when those steps provide the biggest benefit.

New policies and approaches may therefore be needed to ensure convergence and thus achieve real interoperability rather than adding to the issue of divergence.

Federation and Interoperability are issues relevant for many capabilities, but in particular for "Data Management" and "Virtualisation" (Table 3 and Table 6), as well as aspects related to "Cost Reduction" and "Improved Time to Market" (Table 2 and Table 5).

Main issues: proprietary structures / de-facto standards; vendor lock-in.

Virtualisation, Elasticity and Adaptability

Though virtualisation techniques have improved considerably over recent years, additional issues arise with the advent of cloud systems that have not been fully elaborated before – in particular related to the elasticity of the system (horizontal and vertical up- and down-scaling), interoperability and manageability & control of the resources. Changes in the configuration of the service / data need to be reflected by the setup of the underlying resources (according to their capabilities and capacities), but also changes in the infrastructure need to be exploited by the virtual environment without impacting on the hosted capabilities. For example, if another CPU is added to a virtual machine, the running code should make use of the additional resource without having to be restarted or even adapted. This obviously relates to the issue of programming models and resource control (cf. below) – it should be noted in this context that actual resource integration in virtual machines is less an issue than developing applications that actually exploit such dynamic changes.

To provide efficient elasticity that is capable of respecting the QoS and green requirements as listed above, new, advanced scheduling mechanisms are required that also take the multi-tenancy aspect into consideration. For example, it may be more sensible to delay execution if resources will be available shortly, so as to avoid the employment of currently powered-down resources etc.

Virtualisation (and to a degree scheduling) have to take the human factor into consideration thereby: the degree of interaction with cloud systems, as well the increasing connectivity will require that the systems are capable to integrate humans not only as users, but also as an extended resource that can provide services, capabilities and data.

Currently, also little support is available for cross-platform execution and migration which global cloud structures will require (with the exception of specialized "niche" cloud systems). Especially, the movement of (parts of) an application between cloud structures (e.g. from private cloud to public cloud and back) is a key issues that is not supported yet.

All these capabilities will require a stronger "self-*" awareness of the resources and the virtual environment involved, so as to improve the adaptability to changes in the environment and thus maintain boundary conditions (such as QoS and business policies). And, of course, implicitly new models to develop according applications and tools that can easily exploit these features (cf. below).

See in particular Table 1, Table 4, Table 3 and Table 6 for an overview over the respective capabilities and how they are currently addressed.

Main issues: elasticity; optimised scheduling; interoperability; resource manageability; rapidly changing workloads.

APIs, Programming Models & Resource Control

Cloud virtual machines tend to be built for fixed resource environments, thus allowing horizontal scalability (instance replication) better than vertical scalability (changes in the resource structure) – however, future systems will have to show more flexibility with this respect to adapt better to requirements, capabilities and of course green issues. In addition, more fine grained control over e.g. distribution of data etc. must be granted to the developer in order to address legislation issues, but also to exploit specific code requirements.

Cloud systems will thus face similar issues that HPC has faced before with respect to description of connectivity requirements etc., but also to ensure reliability of execution, which is still a major obstacle in distributed systems. At the same time, the model must be simple enough to be employed by average developers and / or business users.

Cloud systems provide enhanced capabilities and features, ranging from dynamically scalable applications and data, over controlled distribution to integration of all types of resources (including humans). In order to exploit these features during development of enhanced applications and services, the according interfaces and features need to be provided in an easy and intuitive fashion for common users, but should also allow for extended control for more advanced users.

In order to facilitate such enhanced control features, the cloud system needs to provide new means to manage resources and infrastructure, potentially taking quality of service, the green agenda and other customer specifications into consideration. This, however, implies that future cloud systems have to discard the classical layered model (see also [29]). Development support for new "cloudified" applications has to ensure movability of application (segments) across the network, enabling a more distributed execution and communication model within and between applications. Since cloud applications are likely to be used by much more tenants and users than non-cloud applications ("long tail"), customizability must be considered from the outset.

The issue applies equally to distributed code, as to distributed data. Data is expected to become exceedingly large (see "Data Management" above) - hence an interesting approach in cloud system's code management consists in moving the software to the data, rather than the other way round, since most code occupies less space than the data they process. However this is intrinsically against the current trend for clouds to be provided in remote data centres with code and data co-existing.

This relates to the issues identified in Table 1, Table 4, Table 3 and Table 6.

Main issues: connectivity; intelligent distribution (code & data); multi-tenancy; enhanced manageability; reliability; ease of use; development and deployment support.

2. Non-Technical Gaps

Legislation, Government & Policies

Not only data (cf. above) is subject to specific legislation issues that may depend on the location they are currently hosted in, but also applications and services, in particular regarding their licensing models. Legislation issues arise due to the fact that different countries put forward different laws regarding which kind of data is allowed, but also which data may be hosted where. With the cloud principally hosting data / code anywhere within the distributed infrastructure, i.e. potentially anywhere in the world, new legislative models have to be initiated, and / or new means to handle legislative constraints during data distribution.

Related to that, governance of clouds needs to be more open to the actual user who needs to be able to specify and enforce his / her requirements better (see also resource control above), such as data privacy issues, issues caused by business (process) requirements and similar. Governance solution could also help to select only those vendors providing open-source solutions, thus avoid vendor lock in.

Clouds generally benefit from the economic globalisation so that providers (and implicitly users) can make use of cheaper resources in other countries etc. Hence, similar issues apply to clouds that apply to the global market and new policies are required to deal with jurisdiction, data sovereignty and support for law enforcement agencies new cross-country regulation have to be enacted etc.

See also Table 3 & Table 6 ("Security", "Data Management", "Multi-Tenancy" etc.), as well as most economical aspects (Table 2, Table 5).

Main issues: legislation; governance; licensing; globalisation.

Economic Concerns

In order to provide a cloud infrastructure, a comparatively high amount of resources needs to be available, which implies a considerable high investment for start-up. As it is almost impossible to estimate the uptake and hence the profit of services offered to the customers, it remains difficult to assess the return of investment and hence the sensible amount of investment to maximise the profit. With the cloud outsourcing principle being comparatively new on the market, new knowledge about business models, market situation, how to extract value and under what conditions etc. are required – in other words, new expert systems and best use recommendations are required.

This also includes issues related to the "green agenda", namely policies basing on dedicated benchmarks under what circumstances to reduce resource usage and / or switch between different power settings etc. This implies new scheduling mechanisms that weigh green vs. business (profit & quality) issues. In a cloud environment it would be possible to improve 'green' credentials by utilising more efficient processors and memory. A few large data centres with clouds are likely to be more 'green' than millions of smaller but already large data centres. Fan et al. argued that up to 50% savings in energy consumption are possible for data warehouses [30]. Notably, from a global perspective, sharing resources may be greener than down-powering idle resources, if this reduces their production (and hence the according carbon footprint) in the first instance.

In general, business control is principally possible, yet linkage between the technical and economical perspective is still weak and hence maintenance of e.g. service quality respecting the economical descriptions still requires improvement.

An indirect economical issue that will have to be solved through e.g., means for improved interoperability (see below), consists in the current tendency towards vendor-lock in. Most vendors want to maintain this status in order to secure their customer base, yet with scope and competition growing in the near future, it is to be expected that even larger vendors will adopt more interoperable approaches. As a side note it should be mentioned that already some major key player are basing their system on more standard based approaches, such as MS Azure.

See also all issues in the economic issues tables (Table 2 & Table 5).

Main issues: extended business knowledge; improved QoS management; Green Agenda; energy proportional computing.

IV. TOWARDS A EUROPEAN VISION

Cloud systems are no pure research aspect, but a commercial reality that fulfils the required capabilities, even though typically only to a limited degree (considering its full diversity). What is more, cloud provisioning is currently predominated by the American market with the first efforts of Europe only slowly arising in commerce. Even though Europe owns a rich set of resource infrastructures, the US have a considerable advantage and employed it to set up various cloud systems (such as Amazon, eBay, Google, Microsoft).

On the other hand, one can note a strong similarity between the business incentives for Grid vendors and Cloud providers, as well as strong overlaps in the technological basis. Indeed many European Grid vendors are already moving their offers from Grid to Cloud concepts, enhancing in particular on the elasticity and pay-per-use aspects. The according hurdle of infrastructure availability and of the technological adaptation process is thus lower for these vendors, providing Europe if not with a head-start, so at least with a good starting position. Obviously, this does not imply that only Grid vendors can become Cloud providers and in fact many Cloud providers already set up their own infrastructures and capabilities independent of any Grid technologies.

Most service providers and data-centres will employ cloud infrastructures for their internal use, but also to support the quality of services and capabilities they sell to the customers. With the varying scope of requirements, including location, legislation and cost, cloud infrastructures cannot be restricted to a single nation or country, but instead will span a global network.

Such a “loose federation of cloud systems”, where a virtual environment can principally be dispersed all over the world, shows strong similarity to the original ideas of distributed computing, utility computing and grid systems.

It is therefore of particular interest to identify how and to what degree especially Europe can contribute to realising this vision. This section will analyse the specific strengths and weaknesses of Europe’s industry and research community to identify the specific opportunities of Europe to shape and participate in the cloud future.

A. SWOT ANALYSIS

The SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis is a means to identify the particular areas where Europe can contribute and even lead the development and uptake of cloud systems in a global market. The following overview highlights the main important aspects that Europe can and should pursue – it is noticeable thereby that Europe’s specific strengths rests on the consolidated effort to address cloud systems on a more global scale than the US can do. This is particularly relevant to enable the global “loose federation of clouds” vision that integrates the control layers into an enhanced resource management and integration model where consumers and both large and small enterprises can equally participate.

Most current approaches towards infrastructure management tend to add further abstraction and manageability layers on top of existing ones, thus complicating the structure and making low-level interoperability on a resource level more complicated. As also identified in the Next Generation Grid report #3, the layered and stacked approach of “classical” middleware approaches is counterproductive to future application needs [29] and hence needs to be re-assessed. A stronger convergence with eInfrastructures [51] is therefore to be expected – see also “Analysis” (section V)

for a detailed analysis. In order to achieve this, more international consolidation approaches will be required to align different end-user positions.

Strengths	Weaknesses
• Knowledge background and expertise in related technological areas • Significant expertise in building high-value industry specific applications • On-going research projects and open source technologies • Strong SOA and distributed systems research community • Strong synergies between research and industry; technological platforms • Concertated government effort (legislation etc.) • Selling products & telecommunications (as opposed to selling new technologies) • Provisioning of complex processes as services, rather than of low level infrastructures • Strong telecommunication industry (research, consumer focus, investment capabilities) • Commercial success-stories	• Few resource infrastructures available in Europe • Comparatively weak development of new (cloud) technologies in comparison to US • Primarily consumer; main Cloud providers are not European • Research timelines vs. fast moving markets • No market ecosystem around European providers • Subsidiaries and fragmentation of key industries • No platform to find / select cloud providers
Opportunities	**Threats**
• Strong experience and involvement in standardisation efforts • European companies use (and need) their own clouds (private clouds) (cf. location) • Growing interest from both industry and academia in cloud technologies (cf. “readiness”) • Existing infrastructures with strong resources and in particular with strong communication networks (e.g. telecoms) • Clouds provide an excellent backend for mobile phone applications (which have usually low power local resources). • Increase competiveness and productivity, of service providers by adoption of local/hybrid/public computing platforms • Application provisioning instead of technology orientation • Support SMEs and start-ups with improved ROI (elasticity), reduced time to market and easy adoption • New business models for cloud improved products and cloud adopters • High awareness for the green agenda and new approaches to reduce the carbon footprint • Similar business incentives and infrastructure requirements between Grid and Cloud, facilitating the movement from Grid to Cloud Provider	• Better developed cloud infrastructures (mainly in the US) already exist • High investment and funding required to build up infrastructure • Investment/economic benefit asymmetry (IPR, OSS, commercialization) • Lacking IaaS provider(s) • Dependency on external (non-European) providers • Technological impact / development underestimated • Latencies (federation too inefficient)

TABLE 7: SWOT OVERVIEW

The following sections provide a more detailed elaboration of the main SWOT aspects.

1. Strengths

Europe has a particularly strong telecommunication industry that can be an important commercial factor for the US to consider in their future cloud related development. Accordingly, Europe *does* have the economic strength to impact on the US.

The main strength and hence advantage of Europe, however, consists in its consolidated and synergetic efforts to address new technological trends and governmental issues – this implies in particular issues related to the interoperability and convergence of technologies, as well as to global policies and legislation approaches. More than the US, Europe has therefore the strength to address control and management aspects related to a global cloud infrastructure. Europe thereby has the specific role as a technological and governmental counsellor / advisor.

As Europe is also very strong in selling products rather than new technologies, it should be examined how cloud capabilities can be exploited to enhance the capabilities and qualities of services and products in the European market. Especially, Europe's strong SOA research community can be exploited to help industry to develop the tools and methods to build cloud applications. Also, most US companies concentrate on the consumer market (and are hence more visible), whereas Europe focuses particularly on provisioning of professional services. Europe would hence act as an adopter of cloud technologies providing and building applications that are used by cloud users world-wide.

Most research projects pursue a strong open source approach, which is beneficial for both the community pursuing existent results further, as well as for uptakers that do not want to be restricted to a specific vendor and / or want to adapt the application / service to their specific needs. It should be noted in this context that Europe has a strong background in open source code development, even though they are mainly exploited through U.S. companies [32] – for example, the well-known open-source virtualisation platform Xen was originally developed under UK research funding of the Engineering and Physical Sciences Research Council (ESPRC) [55]. Nonetheless it can be noted, that even public bodies in Europe are open for employing open source applications [31].

2. Weaknesses

However, Europe is already behind the development in the US and considering the timelines of research to reach market-readiness as opposed to the fast movements in the market itself, time is a critical resource with respect to positioning Europe in the global cloud development.

Along the same line, it is up to investigation in how far the European market, and in particular European providers can be considered "ready" for migration to cloud systems: not only does this entail a change in their current modus operandi, including the actual service logic and code, but also does this require a substantial starting investment in order to gather and prepare the infrastructure.

Accordingly and considering the current situation on the market, European industry has a stronger tendency towards being a cloud consumer or adopter than a real public cloud provider. However, due to the amount of end-users, cloud based applications may find a bigger market than actual cloud infrastructures.

3. Opportunities

In general, gaps identified in section III.C build a basis for adding value to existing cloud infrastructures and / or building new added value cloud services - as such they build general opportunities for cloud related research and development. However, these gaps are not necessarily specific to the European situation.

It has been noted that cloud systems are not restricted to public clouds – instead, most providers will initially want to make use of private clouds and in the long run will employ a hybrid cloud infrastructure so as to address the issues of control versus cost. In combination with the issue of legislation and data distribution, this builds a requirement for European industry to have cloud technologies and infrastructures at their disposal within national boundaries, so as to ensure that data *can* remain within a legislative area, if required.

Related to this, Europe has a wider market and governmental structure at its disposal and accordingly more expertise and influence on global policies, legislation issues and global business models than most other nations. This expertise and capability will prove particularly useful to build up new global policies and regulate cloud specific legislations.

Similarly, this knowledge can be employed to provide the environment into new business models and expertise to ensure economic value creation from the employment of cloud systems for various use cases. This information can be used for new systems that automate the cloud configuration even more efficiently. The issue implicitly relates to aspects of Green IT, which currently has found little support in cloud systems, but is a significant issue in current datacentre design.

It should be noted here that, just because the "cloud" in Europe is not visible, it does not imply that it does not exist: in fact, just like the Grid, plenty European companies already employ cloud technologies for the provisioning of enhanced services to their customers. As noted, the concept of cloud computing is not new as such and as opposed to many other technology, not first driven by research but developed and exploited from *a commercial* perspective from the beginning. Europe hence already has a comparatively strong background in (indirect) cloud provisioning, and its industrial players already show the relevant business incentives to take the final steps towards cloud usage. However, there is little effort being vested into making the according systems publically available, i.e. European vendors typically employ cloud strategies for improved service provisioning (cloud adopters & vendors) rather than selling cloud infrastructures (cloud providers or resellers).

Overall, there is a growing interest in cloud technologies from both industry and academia which provides a specific opportunity for Europe to participate in this global movement.

4. THREATS

These opportunities are obviously counterweighted by some threats that particularly relate to the effort involved in such a participation (see above), namely speed, i.e. the time it takes to address the opportunities versus the market development; and cost, such as starting investment for building up local resource infrastructures etc.

Overall, the US has currently a better developed cloud infrastructure and Europe runs a high risk of becoming dependent on external, i.e. non-European providers, if it only acts as a supporter / counsellor and / or an adopter. It should be noted in this context though that many global providers open (and operate) *datacentres* (and hence potential cloud infrastructures) based in Europe.

In addition to these primarily economic issues, technological threats may pose additional issues, where it comes to overestimating the capabilities of clouds and underestimating the restrictions and challenges. Particular potential threats in this area that can already be identified relate to: latency may prove to have too much impact on distributed (or interactive) computing, thus necessitating better analysis of connection requirements and improved data segmentation / distribution etc.; dynamic systems may impact on speed of distributed systems if too many reconfigurations take

place; future resource sizes and capabilities may make clouds unnecessary; the divergence of future resources becomes unmanageable in a single infrastructure etc.

Please also refer to Appendix A – Other Developments for more details on technological threats.

B. Specific Chances for Europe

Basing on the SWOT analysis above, as well as the identification of gaps and open research issues in section III.C, this section will elaborate the specific main opportunities for Europe in the development of a global cloud ecosystem. It is generally accepted that Europe has the capability and the capacity to join such an ecosystem and would contribute vitally to such a goal. A particular strength of Europe thereby consists on its consolidated and joint efforts in all issues related to research, legislation and (governmental and commercial) policies.

It is also generally acknowledged thereby that in particular the US has an advantage over Europe with respect to the development and provisioning of already existing cloud infrastructures (even though they are mostly still in a beta / testing phase) that show little convergence as such though.

1. Towards Global Cloud Ecosystems

Europe will participate in the movement towards a global cloud ecosystem, due to a growing interest of industry and academia, as well as a specific requirement for location specific resource infrastructures. Such global ecosystems would be useless without the capability to easily switch between providers / resources and without ensuring that specific legislation and policies are met.

Europe, with its specific background in joint research efforts, convergence in legislation and international policies etc. is a key enabler in this vision by keeping the “big picture” in mind when defining cloud behaviour, interfaces etc. (cf. Holistic Systems below).

Similarly, Europe’s cloud computing research agenda could centre its efforts to be the centre of excellence for cloud applications in key business areas for European companies (key industries and SME’s). With a strong focus on the usage patterns and demands from European industry, a “user-driven” research agenda promises to provide significant impact on the economic agenda.

“Globalisation” in this context involves in particular the following issues:

- Global legislation issues
- Strong European partner eco-systems
- Behaviour policies
- Interoperability & standardisation efforts

2. New Business Models and Expert Systems

Extracting value from cloud system employment is not always straight forward, as it depends on the cost and effort to be invested first versus the (potential) gain from the employment of such a system. There is little knowledge so far about when and under what circumstances to move to a (public or private) cloud, respectively when to distribute capabilities in a hybrid cloud.

Though outsourcing to clouds can reduce start up time and makes better use of resources due to the elasticity of the infrastructure, the additional effort to move services and large datasets into a new environment, as well as the risk to lose control over the system, makes such a movement a considerable business decision. As long as interoperability is at a stage where no simple movement from local to cloud platforms is possible (cf. i) and iii)), additional knowledge is required to support

such decisions and in the long run allow for autonomic management of outsourcing and reconfiguration decisions.

Main knowledge to be gained relates to:

- How to create and extract value
- When to outsource (and where to)
- How to improve ROI
- How to reduce start-up time
- How to build cloud-ready applications

3. Holistic Management and Control Systems

Employing cloud systems more than ever requires a holistic view across all horizontal and vertical issues: not only is it necessary to supervise the distribution of services, code and data across the infrastructure (horizontal), but also improved control over the individual middleware and resource layers and communication protocols (vertical) is needed. This is important to address scalability issues, as well as to ensure adaptability to individual requirements. In this context, customizability and multi-tenancy are of importance.

In order to realise such capabilities, new control and management systems are required that integrate the horizontal and vertical view. With the research background in Grid and SOA, as well as the expertise in varying tiers of such infrastructures, the consolidated efforts of European industry and academia can significantly support the development of a holistic, integrated and nation-wide cloud infrastructure (cf. global cloud ecosystem).

New infrastructure models need to:

- Integrate all tiers and layers
- Address cross-boundary scalability, elasticity and multi-tenancy
- Respect policies, legislations and business knowledge
- Manage all aspects related to composition and execution management

4. Cloud Support Tools

Europe can offer new features and capabilities to support cloud employment and to improve adoption (see also mediation of services). Europe can build on its particular knowledge and consolidated research efforts to identify gaps in current provisioning models, as well as to address them by providing supporting tools.

Such tools would cover issues related to:

- Supporting to build up new platforms easily
- New programming models and tools that deal with distribution and control
- Enhanced features for provisioning, including respecting business obligations
- Improved security and data protection
- Efficient data management
- Energy efficiency on all layers
- Easy mash-ups of clouds exposing a single user interface etc.

5. Mediation of Services and Applications on Clouds

A specific strength of Europe consists in selling advanced products and, as such, in aggregating or accumulating existing capabilities to offer enhanced products and services. Related to enhanced

support tools (see above), Europe can exploit the capabilities offered by (existing) cloud systems (cf. section II.B) to enhance the capabilities of products and services offered through European industry. Traditionally, Europe has an excellence in utilizing and benefiting from building high-value European applications on top of global platforms vs. focusing on the underlying platforms itself.

Pure infrastructure and application-services do play a role in the low-end; enterprises however request complete business-processes as a service. And here the service providers in Europe (those originally coming from IT-services as well as those coming originally from the Telco-business) are very active. The infrastructure for these business-services is in many cases already provided as a cloud, specifically when the volume varies dynamically. The services sold however are e.g. called 'dynamic services' or 'Business Flexibility'. The use of cloud in today's professional services is quite high in Europe.

It may be worth noting in this context that already new service providers enter the market that make explicit use of cloud capabilities in order to reduce their cost of investment and improve the availability and reliability of their services [36].

Extended features can thereby include amongst others:

- Improved accessibility & availability
- Scalability according to needs
- Enhanced computational power
- Customizable products
- Composition / aggregation of higher-value products / applications based on existing ones

6. Green IT

Reducing the carbon footprint becomes more and more relevant in industry and IT. Europe has strong expertise in these areas through policy making and extensive research, from which the cloud systems can benefit. Of particular interest in this context is the exact threshold for up- and downscaling in cloud systems, as well as energy proportionality at all levels of the system. But also essential policy measurements are needed to compensate for the additional carbon emission through building up and maintaining cloud infrastructures: due to competition, any energy savings will automatically be invested into new resources so that the net consumption stays the same – energy efficiency alone is hence not sufficient to address the green agenda.

Europe has gathered various experts in this area to develop improved policies and techniques for reducing the energy consumption, which need to be extended to cloud systems. This relates strongly to i) and ii).

7. Commodity and Special Purpose Clouds

A strong adoption opportunity in the current market consists in both commodity and special purpose offerings. Whilst commodity clouds would support the global vision of cloud computing, where platforms offer similar capabilities, special purpose clouds can be seen as (customizable) extensions to commodity clouds that serve the specific needs of individual consumers, e.g. extended data archives with analytics functionalities etc. Future global clouds will allow composition of such enhanced features to cover a broader scope of customers.

Europe provides a wide range of *consolidated* expert groups in different areas which are supported through various infrastructures and collaborative environments. By offering the according capabilities through special purpose extensions to commodity clouds (or special purpose clouds),

experts all over the world would be able to make use of these features in a scalable and hence adjusted to need fashion.

8. Open Source Cloudware

Most available cloud systems these days are provided as closed source or internalized open source, so that the community can contribute little to its development and convergence & interoperability is complicated. In order to ensure convergence, customer driven approaches are needed, which often imply open source solutions – in particular research results should follow the open source initiative to simplify uptake and support convergence.

Open source is thereby not restricted to usage in research communities for publication of project results, but also finds high uptake on the end-user side, as can be seen by web site statistics of open source community sites, such as Sourceforge. But also public and governmental bodies in Europe take up open source solutions for supporting their work.

Europe has large open source communities and a strong background in open source development and provisioning. Nonetheless, as the European Software Strategy industry report [32] [33] indicates, many of the open source technologies developed in Europe are exploited by US companies. According to one estimate, 90% of the business derived from open source systems is generated by non-European players. Furthermore, most consortia managing open source development and marketing are based in the United States and funded by US IT companies (such as Sourceforge and CodePlex).

If the cloud computing research aims at realizing a *sustainable* European economic opportunity as envisioned in i2010, this imbalance needs to be addressed. A thoughtful "utilization" framework, which allows the broadest set of European companies with diverse business models to leverage this asset, could be beneficial.

9. Movement from Grid To Cloud

Even though Europe generally lags behind the US with regards to the industrial cloud movement and even though Europe seems to have less resource infrastructure at hand, there is still a comparatively large group of Grid vendors and uptakers in Europe. Due to the strong similarity in particular between the *business incentives* of Grid vendors and Cloud providers, as well as due to similar requirements towards the infrastructure, it is comparatively easy for current (European) Grid vendors to move towards cloud provisioning (including supporting tools and middleware) and already being undertaken by companies such as GridSystems.

European market players, particularly from the Grid domain, can hence generally be considered "ready" for a movement towards cloud service offering. In order to execute that step, it must become visible to them how a) this can improve their business, b) why any customers would follow this movement and finally c) how this can be implemented and how potential obstacles can be over-come. This relates to all aspects as identified in section III.C, but requires that an according initial movement is provided soon, respectively that awareness of according support improves quickly.

10. Start-up Networks

Cloud computing is useful for early stage start-ups, both as a low cost alternative to the company's internal IT costs as well as for quick prototyping and scalable/flexible novel services.

Today's funding of start-ups is sparser than before the financial downturn, and VCs are moving away from the very early stage start-up funding, leaving the start-ups to incubators and business angel

networks. A trend [45] in the start-up area is that start-ups try to run further on no, or low, external funding. So called “Microstartups” are evolving, based on today's free/low cost services for small companies IT, and cloud computing.

Pilots building start-up networks supported by cloud computing are evolving [46]. In these pilots hands-on cloud computing courses are given to early stage start-ups, who are invited to use cloud resources for prototyping. E.g. winners of entrepreneur challenges (in Estonia) are now given (in addition to prize money) computation time on cloud resources.

These pilots are now part of larger cloud projects [47] and could be the basis to build, from grass-root level, highly competitive and cloud aware companies in Europe.

All above also applies to established companies internal innovation activities.

V. ANALYSIS

Even if considered cynically as 'hype' it is clear Cloud computing will play a large part in the ICT domain over the next 10 years or more. The major reasons are:

1. more and more enterprises look to outsource their IT
2. some businesses require additional capacity temporarily for particular needs
3. exploit cloud systems for experimental purposes thus avoiding disruptions
4. utilise a cloud service as 'neutral territory' for joint enterprise operations
5. business continuity/disaster recovery
6. provide a low-cost entry point into ICT provision for a company

etc.

As discussed, the technological research and development status is not yet sufficient to fulfil all business needs, which would allow broad usage of clouds for purposes such as listed above. Hence, there is a need to continue research and development to which Europe can contribute essentially.

The following sections will provide an analysis basing on the information provided in the preceding chapters, of how Europe can and should participate in this movement and what this means in particular from a research perspective.

A. SPECIFIC OPPORTUNITIES

There are several business opportunities for Europe requiring R&D in both technical aspects (such as service metadata) and non-technical aspects (such as legalities and business models). Basing on the SWOT analysis in section IV.A, we can foresee in particular the following opportunities as relevant for European participation in the cloud movement:

O#1 Infrastructure as a Service Cloud Provisioning: outsourcing infrastructure to reduce management overhead and to decrease cost for acquiring resources in the first instance. IaaS clouds are the most basic and at the same time most essential form of cloud systems, as most other cloud capabilities can be build up on it. But support for IaaS clouds is not only of interest for Europe as it provides the relevant basis, but also because legislative issues are as yet unsolved (see also "consultancy" below), i.e. in-country cloud infrastructures are required so as to address specific business' needs for local systems, that respect legislative and location boundaries.

From Europe's perspective, IaaS provisioning has two main aspects to it, related to actual application on the one hand and research related on the other:

(1) Europe needs to encourage wider uptake and usage of cloud systems both as providers, as well as consumers – even though plenty of European businesses already *use* cloud capabilities either for internal purposes (private clouds) or for outsourcing local services / functions to (mostly US based) cloud providers. However, this does not meet the requirements for "in-country" clouds and many consumers still refrain from outsourcing sensitive data / services outside their country and / or with little control over location. In particular telecommunication providers principally already own the necessary infrastructure and the business model fits with their existing service provisioning.

Main issues: lacking European cloud *providers* (not users); legalistic issues
Assessment: basic technology available, improvements desirable
Expected actors: Telecommunication industry
Main actions: encourage uptake
Timeline to achievement: 1-2 years

(2) Even though the basic technological capabilities for IaaS provisioning are already available, some technological improvements are still needed in particular with respect to resource control and systems management. Current cloud technologies offer little control over the actual resources used, let alone respecting their location, which is a serious obstacle for hosting sensitive code and data. In addition, it is still difficult for cloud providers to adapt their system with individual customer requirements or changes in the existing infrastructure, so that improved support for system management is required.

Main issues: little control over resources and system;
Assessment: basic technology available, manageability and control still weak
Expected actors: all research, particularly telecommunication, distributed systems
Main actions: resource control, systems management
Timeline to achievement: 1-3 years

O#2 Platform as a Service Cloud Provisioning: are essentially task and application area specific development and execution support frameworks and thus required in different flavours (depending on the application domain). Even though most PaaS services are still offered by the USA, their scope is still very limited and platform services such as Google's app engine concentrate on broad, but not very business relevant capabilities. Dedicated platforms would however be very attractive for enterprises to support the development and provisioning of dedicated services to their customers and simplify adaptation to individual needs. It would also allow newcomers in the area to develop and provide new services quicker.

In general, cloud platforms are of global interest and not restricted to the American market: similar to service provisioning, dedicated platforms meeting specific business areas will always be required and will / can grow with the amount of expertise available in the respective field (see also consultancy, below), as well as the extension of capabilities.

A major issue towards broad uptake thereby consists in the interoperability issue faced between different platforms: not only do they build on different platform engines for obvious reasons, but also make use of their individual proprietary data formats. So far, there is no general programming model available that deals with distribution, location and communication, as well as supports the scaling problem both vertically and horizontally, that could be exploited as a basis for development platforms. As noted, also in the context of IaaS provisioning, the scalability and in particular adaptability capabilities of PaaS clouds are thereby still quite limited, too.

Main issues: interoperability; programming models; management and adaptation of the system
Assessment: limited scope of platforms; interoperability problematic
Expected actors: telecommunication and large IT as providers; companies located in Europe as users (platform developers); global consumers
Main actions: encourage provisioning; RTD in distributed system management; expertise gathering; standardisation efforts
Time line: 2-5 years

O#3 Cloud Adopters and Service Vendors (Enhanced Service Provisioning): A specific strength of Europe is and always has been the provisioning of dedicated, enhanced services to various business and users (see SWOT). Though these services are not as visible to the average end-user, such as SAP in comparison, they are nonetheless essential for many industrial areas across the world. As with any dedicated service, any country has a fair chance to be a competitor in this market place and Europe's strong background and expertise in this area serves as an important starting point. Europe

could thereby develop a “free market for IT services” to match those for movement of goods, services, capital, skills.

Not only are new, adapted services always required, new means for combining existing services meaningfully and enhancing available services with cloud capabilities etc. are required to compete on the growing cloud based service provisioning market. Notably, many of these aspects have been and still are subject to various research projects, in particular in the grid domain – however, at the time of writing this, they still have not reached a point where they could be used easily or provide the desired capabilities, let alone meet all the requirements.

In addition to this, current cloud services are still restricted to the environment they run on: once a service exceeds the scope of the infrastructure it’s running on, or requests locations that cannot be served by the cloud system, the requirements cannot be met, leading to failure of the according service / application. What is more, services which actually contribute to steering cloud capabilities across infrastructures will face problems related to interoperability, resource control, systems management etc. Mostly because cloud systems are dealing with a degree of scale and heterogeneity hardly ever faced before.

> *Main issues*: interoperability; programming models; management and adaptation of the system; scalability; heterogeneity
> Expected actors: telecommunication providers to expand their services; any service provider
> *Assessment*: no control over scale, interoperability amiss, heterogeneity is problematic, fragmented base capabilities are available but do not come together or fulfil the required needs
> *Main actions*: build up meta-services, encourage service providers to move to clouds and provide enhanced services, realize cloud mash-ups
> *Time line*: 5+ years

O#4 Cloud Consultancy: the major obstacles towards wide-scope cloud uptake consists mainly in the lack of knowledge about cloud usage, its impact, movement from normal to cloud-based provisioning etc. In particular economical and legalistic issues are still completely vague. This is mainly due to the fact that clouds as a “public” infrastructure are comparatively new in the market and little experience is as yet available about the long term impact from usage and / or about the full scope of usage.

For example, there is little knowledge as yet available about when it is advisable for a service provider to migrate existing services into a cloud environment, let alone, how to execute this migration, i.e. how much effort is worth vesting into such a migration. In other words, means to identify services commercially valuable enough to invest the effort into their conversion, as well as how to approach this conversion. Along a completely different track, there are plenty unsolved legalistic issues yet to be addressed, in particular related to the location of data and / or code: most data owners have specific restrictions about the legal boundaries in which their data is hosted and thus refrain from putting it into a cloud environment, where the data may potentially move to countries with a different legislation. These and further issues have direct impact on other research topics, such as that more control over resource location is required in order to address the legislation boundary issue etc.

Europe, with its basically united approach in legislation, but also in market control and a strong research community can play a major role in providing essential consultancy support ranging from active advisory over toolsets, knowledge bases and migration support to the suggestion of new legislative policies.

Main issues: lack of knowledge and experience; lacking expertise; no consolidated legislation and policy building efforts
Assessment: little experience available; most cloud infrastructures come to existence in a trial & error way – makes new providers sceptic...
Expected actors: legal experts, business consultants...
Main actions: analyse the legislative system; analyse the technological and economical basis; gather knowledge and test models; build up an expert system etc.
Time line: 3-10 years

O#1.1 IaaS Provisioning				
	Main issues: lacking European cloud providers (not users); legalistic issues			
	Assessment: basic technology available; improvements desirable	Expected actors: Telecommunication industry	Main actions: encourage uptake	Timeline: 1-2 years
O#1.2 IaaS Technologies				
	Main issues: little control over resources and system;			
	Assessment: basic technology available manageability and control still weak	Expected actors: all research; telecommunication; distributed systems;	Main actions: resource control systems management	Timeline: 1-3 years
O#2 PaaS Technologies				
	Main issues: interoperability; programming models; management and adaptation of the system			
	Assessment: limited scope of platforms; interoperability problematic	Expected actors: telecommunication & large IT; European companies; global consumers	Main actions: encourage provisioning; RTD in distributed system mgmt.	Timeline: 2-5 years
O#3 Enhanced Service Provisioning, Meta-services				
	Main issues: interoperability; programming models; management and adaptation of the system; scalability; heterogeneity			
	Assessment: fragmented base capabilities are available; scale, heterogeneity and interoperability problematic;	Expected actors: telecommunication to expand services; any service provider	Main actions: build up enhanced & meta-services; encourage movement to clouds; realize cloud mash-ups	Timeline: 5+ years
O#4 Cloud Consultancy				
	Main issues: lack of knowledge and experience; lacking expertise; no consolidated legislation and policy building efforts			
	Assessment: little experience available; cloud infrastructures still in experimental stage	Expected actors: legal experts; business consultants	Main actions: analyse the legislative system; analyse the economical basis; build up an expert system etc.	Timeline: 3-10 years

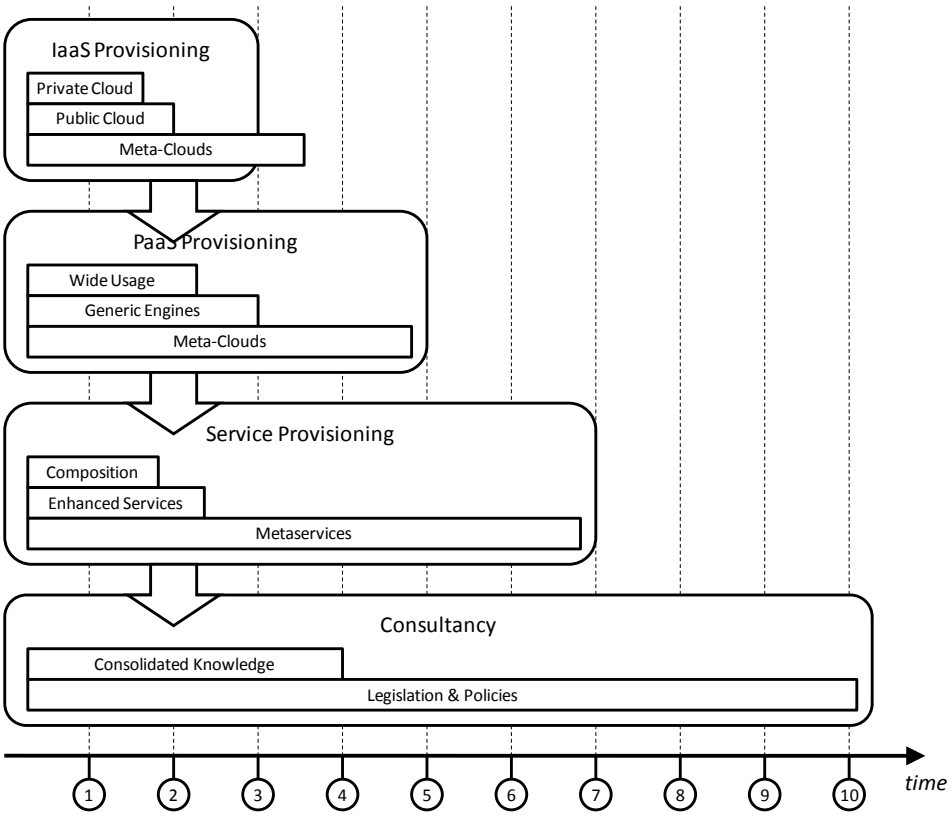

FIGURE 2: ESTIMATED TIMELINES FOR THE INDIVIDUAL OPPORTUNITIES TO REACH THE MATURITY SPECIFIED IN THE REQUIREMENTS. THE TIME AXIS SHOWS THE EXPECTED AMOUNT OF YEARS TO COMPLETION (CIRCLES).

Figure 2 provides an overview over the specific European opportunities and in which time they are expected to reach essential maturity with respect to the capabilities identified in section II.B. Note that obviously all opportunities may be constantly enhanced with according increments in efficiency, resource usage etc. – this report focuses primarily on essential capabilities related with the requirements and capabilities as identified in preceding sections though.

B. RELEVANT RESEARCH AND TIMING

Cloud computing poses a variety of challenges to conventional advanced ICT. Basing on the gap analysis in section III.C and the specific opportunities as identified in the preceding section V.A, we can clearly identify the relevant topics and issues that require further elaboration through dedicated research and development:

1. R&D TOPICS

One can distinguish in particular between technical (cf. section III.C.1) and non-technical (cf. section III.C.2) aspects relevant to meeting the opportunities – the following section explains how the respective topics contribute to addressing the requirements:

Technical Topics

Current advanced ICT solutions are insufficient to meet the technical requirements put forward to cloud systems, in particular regarding the unprecedented scale and heterogeneity of the required infrastructure:

T#1 Scale and Elastic Scalability are considered essential capabilities of all cloud systems (cf. section II.B) but are not even supported to their full degree in most existing infrastructures: neither code nor data are currently structured in a fashion that allows controlling their scaling behaviour efficiently. Most cloud systems achieve scalability through horizontal replication, rather than actually increasing the availability of necessary segments or increasing the resources for specific (sub)tasks only; also, rapid and efficient scale down (destruction of instances) is still a technological problem. Along with this issue comes the problem that the effective usage and needs of applications / users cannot be predicted so as to cater for timely and efficient adaptations.

This implicitly means that *resources are still wasted* unnecessarily and that uptake for both potential providers and customers is still unattractive – in particular in large scale situations, as it may lead to undesired resource consumption. Considering the lack of business expertise and experience in this area (see Consultancy), knowledge about which application / service types behave how in the cloud and hence are most well suited for this type of provisioning.

In order to improve scaling and distribution behaviour, the actual structure of cloud based programs and data needs to be improved through new segmentation concepts and distributed programming models. Communication, latency, user location, and in particular consistency handling will play major roles in this context (see Programming Models) so as to enable large scale efficient applications and thus to pave the way towards meta-services.

Relevant for: O#1.2, O#2 and in particular O#3
Primarily relates to: Virtualisation, Elasticity and Adaptability (p. 32)
Time to finalisation: 5+ years

T#2 Trust, Security and Privacy are on-going research issues in *any* development, as new security holes will appear with hackers advancing in their efforts. In particular in cloud infrastructures, additional issues arise that can be considered serious security and privacy concerns:

First of all and most obvious, direct concerns arise from aspects such as lacking control over data and code distribution in potentially globally distributed infrastructures, security holes in remote servers, potential data loss (as happened to T-Mobile's and Microsoft's Sidekick [53]) etc. Severe security and privacy issues also arise from the fact that clouds provide for multi-tenancy, which needs to be covered full range from shared-nothing to sharing under security constraints. For similar reasons, good provenance mechanisms are needed etc.

Second, and more complicated, indirect issues arise from providing a principally unlimited amount of computational resources to potentially malevolent, respectively untrustworthy entities that may misuse the infrastructure for extreme hacking or denial of service attacks, but also to perform calculations that exceed the current capabilities for average desktop PCs, such as nuclear fusion calculations, if the full potential of clouds is harnessed (cf. T#1). Preventing indirect security threads is obviously even more difficult than addressing direct ones, as their identification requires knowledge about the processes running on the system.

Many of these aspects are related to the lack of a clear legislation model regarding jurisdiction over the hosted data, its distribution in other countries etc. (cf. NT#2). There is a built-in tension between legal and technical availability data placement concerns.

Relevant for: all, but in particular O#1.2, O#2
Primarily relates to: Privacy & Security (p. 30); Federation & Interoperability (p. 31)
Time to finalisation: on-going

T#3 Data Handling: Data size and diversity grows, but current cloud systems are typically restricted either to small data sets (such as profile information) which can be easily replicated or large data sets which are only read. Generally, no support for update-intensive applications or advanced analytic capabilities is offered. Consistency and integrity of the data sets is easily lost due to the concurrent access and wide duplication of data and the lack of provenance makes it difficult to track errors, security issues etc. Cloud systems are also restricted to data-at-rest management and do not allow for e.g. management and usage of streams, unless they are part of the cloud system itself or actually managed via the hosted image.

Clouds exacerbate the known problems of incomplete and uncertain data. With the increased scale and heterogeneity inherent in clouds, the combinatorial effect of incomplete or inconsistent data leads to poor decision-making due to lack of correct or coherent information. Finally, with data stored on multiple clouds and the need to bring heterogeneous distributed data together for various purposes the need for federation of cloud data sources (and matching federation of software) arises. Hence new models, methods and solutions for federating data (moving data to code) and federating software (moving software to data) are needed (see also T#1 & T#4).

Relevant for: O#1.2 and O#2, some O#3
Primarily relates to: Data Management (p. 29)
Time to finalisation: 3 years

T#4 Programming Models and Resource Control: Development on clouds should be simple and intuitive (see PaaS) – however, at the same time the developer will want to be able to control behaviour and location of his application etc. Current programming models offer very little support for scalability (both horizontal and vertical) – in particular in large scale and heterogeneous environments. Parallel applications on the level of meta-services, applications on meta-clouds etc. pose additional issues due to location, distribution, latency, resource control, vertical scale etc. Programming models need to be established to provide sufficient information to programmers to be able to reason about their application designs and their deployment on the cloud without unduly exposing the underlying complexity. At the same time, the model must support manageability of the devised applications and services in a way that allows efficient controlling over distribution and enforcing of resource consumption restrictions on the system side (see also T#5).

To support uptake of clouds, not only new applications and services are of interest, though, but also the migration of existing applications and services to cloud infrastructures. Accordingly knowledge (cf. NT#1) and tools are required to support the migration process, simulating different options and quantitatively reason about behavioural properties of distributed systems.

Relevant for: O#2 partially, mainly O#3
Primarily relates to: APIs, Programming Models & Resource Control (p. 33)
Time to finalisation: 5+ years

T#5 Systems Development and Systems Management: scale and heterogeneity of (cloud) infrastructures grow beyond the point of human system administration and far beyond the point of current system management tools, in particular if specific divergent requirements between resource setups need to be met. Automation of system administration thus requires intelligent capabilities to weigh between requirements and decide on basis of technological and non-technological concerns. Additional capabilities are needed to describe services and allow self-* activity, methods and models for managing dynamic composition, the management of execution within service level agreements, quality of service criteria and criteria relating to trust, security, privacy and cost.

While it is true that most of these areas have been actively researched for decades, the emergence of the Cloud paradigm demands solutions beyond those produced to date in these areas. As mentioned, in particular scalability and heterogeneity pose complete new issues, but also the cloud-implicit problems of latency, distribution and segmentation enhance the problem scope significantly. In particular, the networking and storage components that hitherto were often ignored need to be integral part of the management and design time stacks.

Relevant for: particularly O#2 and O#3
Primarily relates to: Manageability and Self-* (p. 29)
Time to finalisation: 3+ years

Non-Technical Topics

Cloud computing 'asks the questions' of current and emerging business models and legalistics surrounding ICT provision and use. There is a need for research into business models and legal frameworks that – if provided – would assist Europe – and especially SMEs – to overcome the barriers to the provision of and utilisation of Cloud computing.

NT#1 Economical Aspects of cloud systems are still mostly unknown to most providers and users (see also O#4): the usage of and expectations towards scalable systems used concurrently under varying conditions are difficult to estimate and little long-term experience in this direction exists as yet. Even though there is general acknowledgement that clouds can reduce entry time and infrastructure costs for new business entities, there is still little knowledge to support the decisions of either customers (when to switch to a cloud, how much effort to vest into the migration, which type of services are most promising, which cost / infrastructure model works best etc.) or provider (how much does cloud provisioning cost, which kind of scalability and management support works best, which quality of service can be maintained etc.).

Such knowledge is vital however to increase uptake, but also to improve manageability of the system, increase its efficiency, support migration and to improve scalability (cf. T#1, T#4, T#5).

In addition to this, cloud computing offers possibilities to reduce carbon emission through more efficient resource usage – however, this needs to be counterweighed with the indirect carbon footprint arising from a) more experimental (and thus more overall) usage and b) the pressure on cloud providers to update their infrastructure regularly and faster than the average user. The respective concern poses issue on technology (see T#1 scaling) and requires additional economical expert knowledge to be considered in decisions such as listed above.

Relevant for: partially O#2 and O#3, but mostly O#1.1 and O#4
Primarily relates to: Legislation, Government & Policies (p. 33)
Time to finalisation: 3+ years

NT#2 Legalistic Issues: the internet in general is subject to many unclear legalistic regulations, mostly due to the fact that global access is granted from anywhere to anywhere. Similarly, cloud systems typically incorporate resources from all over the world offering them globally to their consumers – with the flexible scaling behaviour of the infrastructures, the location of code and / or data is difficult to control in particular in current infrastructures (cf. T#4). Accordingly, new legalistic arise with respect to which jurisdiction applies, who is liable etc. But not only location poses issues, but also scalability, e.g. in the context of replicating protected code and / or data, i.e. license right and IPR management.

These issues need to be addressed in order to enable clouds on a global (or at least international) scope.

Relevant for: all opportunities, in particular O#4
Primarily relates to: Legislation, Government & Policies (p. 33)
Time to finalisation: 5+ years

Overview

T#1 Scale and Elastic Scalability			
	Relevant for opportunities: O#1.2 – improving efficiency O#2 – partial instead of full distribution O#3 – meta-scale	Relates to gaps: Virtualisation, Elasticity and Adaptability (p. 32)	Timeline: 5+ years
T#2 Trust, Security and Privacy			
	Relevant for opportunities: O#1.2 – improving base security O#2 – federation, multi-tenancy	Relates to gaps: Privacy & Security (p. 30); Federation & Interoperability (p. 31)	Timeline: on-going
T#3 Data Handling			
	Relevant for opportunities: O#1.2 – improved efficiency O#2 – partial instead of full distribution	Relates to gaps: Data Management (p. 29)	Timeline: 3 years
T#4 Programming Models and Resource Control			
	Relevant for opportunities: O#3 – meta-scalable applications	Relates to gaps: APIs, Programming Models & Resource Control (p. 33)	Timeline: 5+ years
T#5 Systems Development and Management			
	Relevant for opportunities: O#2, O#3 –manageability to scale	Relates to gaps: Manageability and Self-* (p. 29)	Timeline: 3+ years
NT#1 Economical Aspects			
	Relevant for opportunities: All – improved efficiency O#1.1 – encourage uptake O#4 – improve business	Relates to gaps: Legislation, Government & Policies (p. 33)	Timeline: 3+ years
NT#2 Legalistic Issues			
	Relevant for opportunities: All – clarify legal issues	Relates to gaps: Legislation, Government & Policies (p. 33)	Timeline: 5+ years

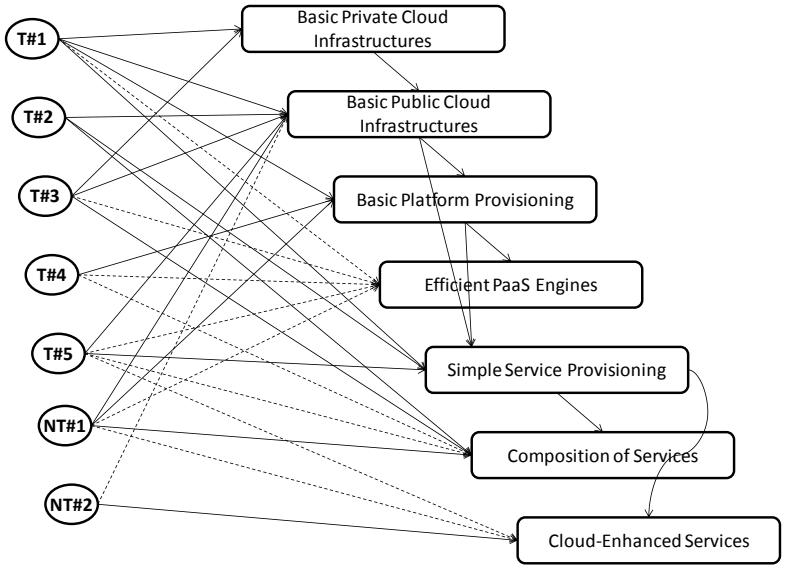

FIGURE 3: DEPENDENCIES BETWEEN OPEN TOPICS AND THE BASIC SPECIFIC OPPORTUNITIES. DASHED ARROWS DENOTE WHERE CURRENT TECHNOLOGIES ARE INSUFFICIENT FOR THE NEEDS OF THE OPPORTUNITY. SCENARIO-LIKE OPPORTUNITIES INHERIT THE TECHNOLOGICAL CAPABILITIES OF THEIR PREDECESSORS.

Figure 3 depicts the relationships between the research topics and the base (scenario-like) specific opportunities which can already be realized today and which make use of the according technological and non-technological aspects. Dotted arrows indicate gaps between the current capabilities of the respective topic and the requirements put forward by the respective opportunity, whilst closed-line arrows indicate that the state of the art technology is sufficient for the direct requirements and can be used for the according purposes; they also indicate which opportunities inherit technological bases from one another. As such, e.g. public cloud provisioning bases on private cloud technologies, but requires additional capabilities in the area of scalability, as the applications running on public clouds are not known in advance as opposed to private cloud infrastructure; in order to fully support all user requirements, public clouds will also require that legalistic issues, such as data location is addressed.

The relationships implicitly relate to the timeline of the opportunities as depicted in Figure 2.

2. PRIORITIZATION

Obviously and as indicated in the text, these topics are of different complexity and even partially depend on one another – as such, e.g. efficient scalability of applications (i.e. segmentation, distribution and replication of code segments) depends on an efficient programming model that enables such behaviour of programs in the first instance and so on. Basing on the research gaps and their relationship to the relevant opportunities, as detailed in the preceding section, one can identify the dependencies between the research topics as depicted in Figure 4.

Realisation of these topics is directly steered and related to fulfilling the specific opportunities (section V.A) which cannot be addressed by the currently available technologies (cf. section V.B.1). It will be noted from the figure that not all developments directly contribute to the specific opportunity – this is either due to the fact that the technologies contribute indirectly (via other developments, such as security via Systems Management) or that the respective aspect forms an orthogonal issue to the respective opportunities (such as legalistic issues which affect all models equally).

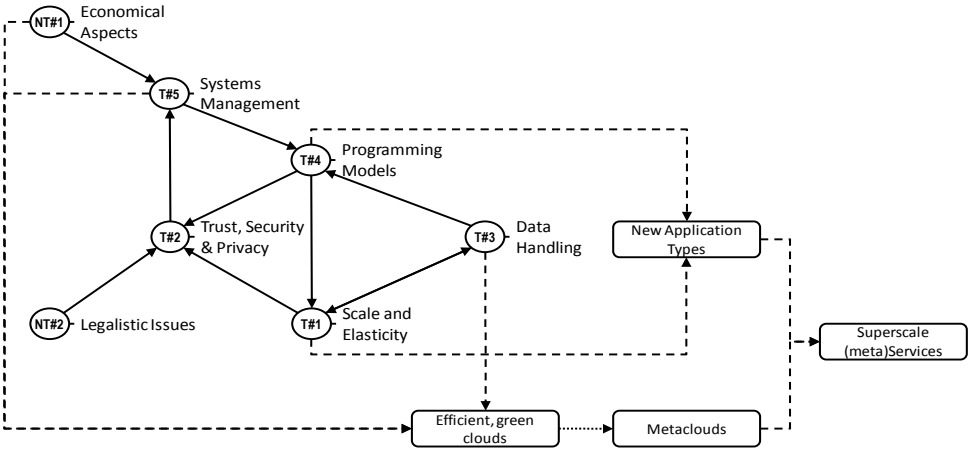

FIGURE 4: DEPENDENCIES BETWEEN OPEN TOPICS AND THE BASIC OPPORTUNITIES. DOTTED LINES INDICATE HOW RESULTS ARE CARRIED OVER INTO SPECIFIC OPPORTUNITIES.

Basing on the dependency analysis and estimated research duration it is therefore possible to prioritize the research topics so as to ensure that the specific opportunities are realized efficiently. Accordingly, the prioritized list may look as follows (cf. Figure 5):

On-going: NT#2 Legalistic Issues

Aspects related to legalistic concerns and policy models will not be solved within the next few years, but will impact on all aspects related to management and provisioning of services, such as data and code location etc. It is therefore an on-going concern that should be addressed immediately.

Priority 1: NT#1 Economical Aspects

Similarly to legalistic issues, gathering economical knowledge is a pressing concern that will be necessary for automated control, as well as to encourage uptake and support usage of cloud systems.

Priority 2: T#5 Systems Management

In order to realize efficient clouds that can handle scalability, elasticity etc. and can adapt according to need, the system needs to be able to be controlled and managed. Essential progress has been made with this respect but needs to be improved to deal with the scope of scale and heterogeneity.

Priority 3: T#3 Data Handling

Similarly, data management is fairly advanced, but is not efficient enough to deal with the data size to be expected in the future – semantic annotation, location and consistency maintenance are thereby considered essential aspects. New segmentation and data analysis / distribution mechanisms need therefore be addressed quickly, before turning towards general efficiency increasing issues.

Priority 4: T#4 Programming Models

Just like data, code is not efficiently segmented, distributed, let alone parallelized – in order to realize future types of applications easily with high efficiency, improved programming models lending from distributed paradigms are required. They will make use of systems management and data handling routines.

Priority 5: T#1 Scale & Elasticity

Improving the scaling efficiency will be an on-going topic in cloud systems but can only be effectively improved once the code and application show better scalability and the system's manageability has been enhanced accordingly.

On-going: T#2 Trust, Security & Privacy

A never-ending issue, in particular in the context of business provisioning, will always be security issues related to authentication, encryption – in particular with respect to issues arising from multi-tenancy and concurrency issues. Notably, scale and distribution pose additional concerns.

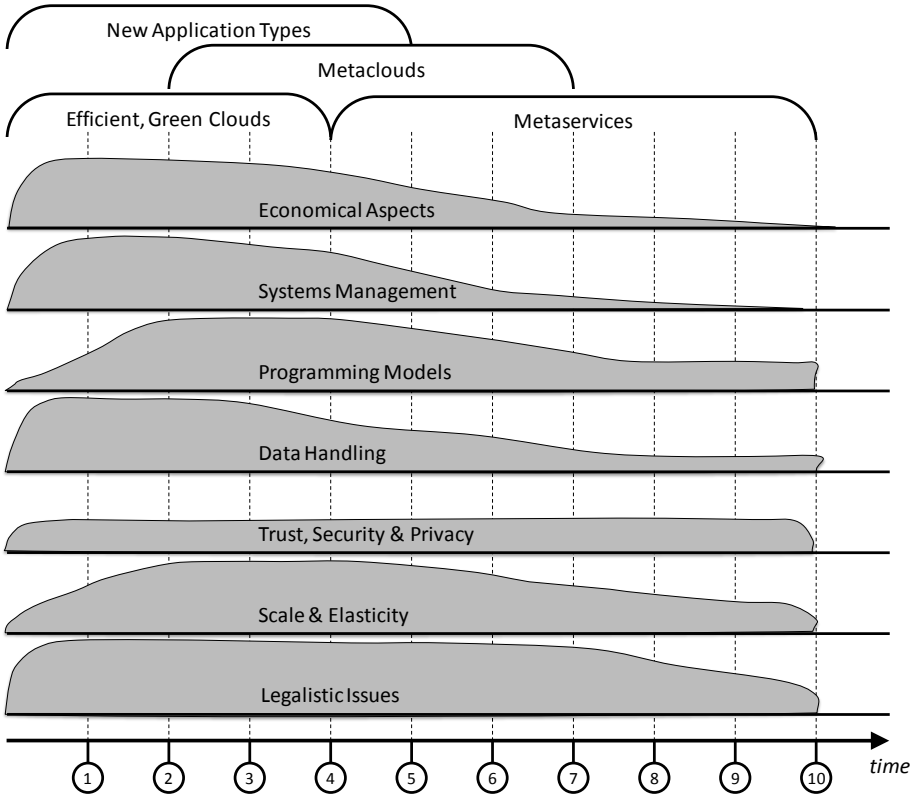

FIGURE 5: RESEARCH TIMELINE (IN YEARS) OF THE INDIVIDUAL TOPICS.
THE HEIGHT DENOTES THE COMPLEXITY / INTENSITY OF RESEARCH TO BE EXPECTED.

C. GENERAL RECOMMENDATIONS

Whatever the view taken by businesses or academia it is clear that Cloud computing in the widest sense presents business opportunities and that to have credible offerings for those opportunities European industry and academia needs to work together to develop the appropriate technologies and other aspects such as economic models and legalistic structures.

The analysis above identifies clearly both opportunities and technical and non-technical topics needed to realize those opportunities. The prioritization is based on the perceived importance of the opportunity in economic terms and the estimated time necessary for R&D on the listed topics to produce useable results.

The expert group recommends that the EC – within the framework programme – opens a special topic on Cloud Computing with the list of topics above as specific work programme elements. In general, STREPs are needed to accomplish the necessary R&D; and the overall architectural and

integration activity requires IPs. There is a clear need for a NoE 'CoreCloud' analogous to CoreGRID to optimize the benefits for Europe form a community of expertise.

D. CONCLUSIONS

The barriers to entry for ICT SMEs concern (a) lack of standardisation of interfaces to guarantee a large 'home market' across the heterogeneity of Europe (b) the heterogeneity of legislation across Europe covering security, privacy, trust, digital rights; (c) the lack of currently long-term-sustainable business models. The barriers to take-up by other business for business benefit include the above but also unfamiliarity with modern ICT and resistance to changing business models.

Clouds offer the opportunity to build data observatories with data, software and expertise together to solve problems such as those associated with economic modelling, climate change, terrorism, healthcare and epidemics etc. Clouds could assist greatly in the e-government agenda by providing information in one place to the citizen, together with software to manipulate the data.

It has been claimed – and indeed demonstrated – that Cloud computing is a green option. Development of Cloud computing in Europe will contribute to reduction in carbon emissions and assist in achieving European targets.

Europe is well-placed to embrace these opportunities due to the excellent background research and development in many of the key technologies such as those associated with GRIDs and SOA (Service Oriented Architecture). However, the provision of an open market in clouds for Europe requires further R&D building upon this substructure. Success will come by intersecting the R&D results with the emerging market opportunities beyond today's Clouds.

Despite the apparent US lead on Clouds there is time for Europe to develop distinctive offerings in several areas, based on well-documented European strengths in ICT. However, this can only be achieved with (a) further technical R&D building upon the success-base from previous framework programmes and national programmes in GRIDs, SOA and other technologies; (b) further R&D on legalistic and business models to find means to lower the threshold barrier for marketplace entry especially or SMEs. Investment in R&D on Clouds brings benefits to the ICT industry, to other industry and commerce, to the media industry, to government and to the citizen. It also offers a greener option for ICT.

APPENDIX A – OTHER DEVELOPMENTS

Cloud computing addresses issues that relate strongly to other research and development areas, as already shown in section II.C. Due to overlap with many existing technologies, developments under way in these areas may have an impact on future cloud provisioning systems. As this exceeds the scope of this report, we will only indicate the most prominent areas in the following:

1. HIGH PERFORMANCE COMPUTING (HPC)

HPC has been dealing with resource pooling and code distribution, reliable execution etc. for a long time now. Though clouds and HPC act on different levels (HPC nodes being more tightly coupled than resources in the cloud), and integrating HPC resources into clouds may not be sensible, there is still a strong overlap between capabilities and boundary conditions that have been investigated in HPC for decades now. This relates in particular to aspects on scheduling, code & data distribution and communication, as well as reliable execution – all issues particularly relevant for distributed, virtual (and dynamic) resource platforms as exposed by the cloud. Depending on the problem domain [50] cloud computing could provide efficient, but also economic viable HPC platforms (example: off peak hours computation and data manipulation vs. guaranteed QoS).

It may be worth noting in this context, that there is a steady movement from HPC technologies to common server machines and even end-user desktops, which may impact on cloud systems in so far, as that they have to cater for complete new resources and hence new management models.

2. BUSINESS PROCESS MANAGEMENT (BPM)

The role of Business Process Management (BPM) technology will increase significantly with the omnipresence of clouds. First of all, the huge number of services available in the cloud will enable a fast and easy creation of new higher-level services by composing the available services. Secondly, the ubiquitous access to application functionality will result in the formation of networks between partners to create competitive advantage by establishing cross-partner business processes.

Cloud technology will significantly ease both, the offering as well as the use of services available. As a consequence, a huge number of services will be available in the cloud and these services will be composed into new services. These services may become available on the cloud again (Composite as a Service) further increasing the number of services in the cloud. The composition of services into new services is supported by orchestration technology. Orchestrations are typically defined by domain experts with some level of IT skill. Supporting a much broader community in composing new services, easier and domain-specific languages for orchestrations have to be provided.

The availability of cheap services providing broad application functionality to everybody implies that companies can no longer distinguish themselves by the use of such (formerly expensive) application functions. One way to distinguish oneself will be the cooperation with partners by establishing a partner network. Many such cooperations will be defined by means of choreography technology reflecting the partner networks. Such choreographies define cross-partner business processes defining very complex and optimized interactions between the partners. The business processes describing the local partner behaviour will be hosted and run in the cloud, being integrated into a choreography. The competitive advantage of a partner network will be monitored and analysed continuously and adapted if needed by exchanging individual partners and the representing choreography itself.

APPENDIX B – (BUSINESS) SCENARIOS

Cloud systems find a wide range of application in varying scenarios – the most promising of them have already been outlined in section V.A. In this chapter we will examine these (business) scenarios in more detail in so far as they may serve as a "guideline" for future application of cloud technologies and thus implicitly as a reference for how the technological gaps may be employed in real business cases.

1. WEB MEGASERVICES

Megaservices act on top of existing services and platforms, combining and extending them so as to provide new, enhanced capabilities. Cloud infrastructures thereby play a secondary, supporting role, focused in particular on the large scale of such services with respect to the amount of underlying instances and resources it has to handle (related to vertical scale), as well as the potential number of concurrent accesses and usages (section III.C.1 "Virtualisation, Elasticity and Adaptability").

Examples of existing megaservices are on the one hand large search engines acting across a large amount of resources (Google Search, MS Bing etc.), and social network sites integrating media and different service types (Facebook, StudiVZ etc.).

In such cases, cloud infrastructures do not only enable easier start-ups for providers with lacking resources to deal with the scale of usage, but what is more can offer integrating support across existing cloud provided services (section III.C.1 "Federation & Interoperability").

The key business benefit is in providing 'mashed-up' novel information for example location of utility paths (cables, pipes etc) under roads placed on a geographical map / image where great savings can be made in minimising the digging-up required to locate faults. Similarly management information of sales by region, distances of supply lines (both for manufacturing and military purposes) is made more understandable.

2. ESCIENCE/EENGINEERING

Traditionally a High Performance Computing (HPC) domain, eScience and eEngineering have high computational demands in order to execute their calculations. Nonetheless, most applications actually do not require full HPC support, i.e. do not execute parallelized tasks, but "only" multiple tasks in parallel and are therefore closer to P2P computing (such as BOINC) than HPC and are most often developed on Grid platforms. In both cases, development of the according applications that allow for distributed (optimally parallel or coupled) execution is typically more complex than an eScientist and / or an eEingineer wants or should have to deal with.

The particular benefit of cloud systems are (1) their ease of access and usage, and (2) their scalability. In particular with parallel task execution, cloud infrastructures can offer horizontal scale up and down according to the respective application's needs (section III.C.1 "Virtualisation, Elasticity and Adaptability") and additional requirements as specified by the user, such as cost restrictions (section III.C.2). As for parallelised tasks, future programming models (section III.C.1 "APIs, Programming Models & Resource Control") will have to enable vertical scale out in an easier fashion, thus making better use of the available resources in the cloud.

At the same time, with cloud infrastructures being easier to set up, development can start locally and extend to external resources on demand, thus relieving the user from having to deal with deployment and connectivity issues (section III.C.1 "Manageability and Self-*" as well as "Federation & Interoperability").

There exist always problems that require the massive power of linked computers to collate and manage heterogeneous information and perform analysis and simulations. When these two aspects are interlinked a virtuous circle of increased scientific understanding is achieved. This has great value in improving the quality of life (e.g. climate change, environmental management, epi-demiology) but also commercial e.g. drug effect simulation or complex engineering assembly design.

3. Traditional IT replacement

The concept of thin clients found a growing popularity in the 1990s as a means to replace expensive local desktop computers with high power servers and multiple access terminals that were comparatively cheap and incorporated little performance capabilities. Web based applications follow the same principle and obviously cloud infrastructures offer the possibility of easy cloud outsourcing, even though the point at which outsourcing becomes economically beneficial may not always be known (section III.C.2 “Economic Concerns”).

Notably, cloud based IT outsourcing covers the whole range from resource infrastructure to complex services / applications hosted on remote machines. Along the same line, it covers the full range of security and privacy concerns (section III.C.1 “Privacy & Security”), as well as data management (section III.C.1 “Data Management”) and federation issues (section III.C.1 “Federation & Interoperability”).

As resources become cheaper and more powerful, most business entities already own infrastructures that can be employed for basic service provisioning, ideally supported with the dynamic self-managed elasticity of private cloud systems (sections III.C.1 “Manageability and Self-*”, “Virtualisation, Elasticity and Adaptability”, “APIs, Programming Models & Resource Control”). Only with growing demand and / or with more relevant services being executed in the local infrastructure, other infrastructures (such as public clouds) should add to the local capabilities (sections III.C.1 “Federation & Interoperability”). Obviously, this implies that all legalistic (section III.C.2 “Legislation, Government & Policies”) and economic (section III.C.2 “Economic Concerns”) of the respective provider are respected.

There are two business-based scenario classes related to this aspect. A company may decide to concentrate on its core (non-IT) business and outsource IT using Clouds and IaaS. This business scenario effectively transfers the investment in-house to a less expensive investment externally. Alternatively the business may decide to use Cloud services to provide business continuity / disaster recovery. An immense business value can (only) be realised if the service is used.

4. Internet of Services

In the generalized Internet of Services vision, services get repurposed, composed, brokered and re-channelled, such as in the context of Virtual Organisations, distributed workflow execution etc. Typically, such composition requires an additional computing layer on top the base service provisioning to enable tasks such as discovery, mediation, brokerage, monitoring etc. so that one can actually talk of two resource levels, similar to the megaservices mentioned above.

Both levels can actually be supported through the scaling and dynamic capabilities of cloud systems, but it will be noted that different requirements with respect to scalability, availability and location apply to these levels. Accordingly, requirements and restrictions from all these areas should be easily configurable (section III.C).

The business benefit is in reduced software development costs (re-use, repurposing), increased software reliability and reduced maintenance costs (previously well-used code re-used), flexibility

(plug-and-play services) providing business opportunities and IT support of them with reduced costs. Within a Cloud environment the service metadata and interfaces are somewhat standardised (although it may be proprietary standards) to realise these benefits.

5. Internet of Things

As already noted in section II.C.2, whilst the clouds do not directly integrate / relate to "things", they can nonetheless offer valuable support for the Internet of Things to support dealing with large, dynamic and distributed data sets. The principles of cloud systems to enable dynamic scale, routing and virtualisation technologies would be particularly beneficial for complex event, data and stream processing between, from and to devices.

In order to enable cloud platforms to participate in the Internet of Things settings and offer support for the complex, potentially location dependent services (section III.C.1 "APIs, Programming Models & Resource Control"), the typically request-response like data transaction behaviour of cloud systems need to be extended (section III.C.1 "Data Management").

An internet of things composed of many detectors and services to manage them has the characteristic of rapidly varying data volumes and rates. Clouds provide an elastic facility to manage this variability. Of course a Cloud environment can also provide the services for analysis of the data streams often associated with synchronous simulation to aid the provision of information to the end-user in an optimal form. The business benefit occurs in applications such as environmental monitoring, healthcare monitoring where the high volumes and rates of data need rapid processing to information for understanding. However, any control system has these characteristics whether the system is for energy (control of power stations), transport (e.g. rail network) or production (production line).

6. Real-time Services

Business environments which depend on real time service provisioning / computation could benefit greatly from the dynamic distribution (section III.C.1 "Virtualisation, Elasticity and Adaptability") and location control (section III.C.1 "APIs, Programming Models & Resource Control") possible in globally distributed cloud infrastructures (section III.C.1 "Federation & Interoperability"). In such environments, latency and availability / accessibility play major role in fulfilling real time requirements and accordingly need to be respected both by the service itself, as well as the hosting infrastructure (i.e. the cloud system).

Environments which have to fulfil real-time requirements often pose specific privacy (section III.C.1 "Privacy & Security") and regulatory (section III.C.2 "Legislation, Government & Policies") requirements towards the infrastructure, due to the competitive nature in this space. Implicitly, most infrastructures will tend to be private or have to observe special purpose regulations.

The business benefit is found in the ability to manage real-time external events with the Cloud environment being sufficiently responsive and elastic to 'keep on top' of the external situation. This aspect links closely with scenario 5above, but emphasises the need for real-time monitoring and control for applications particularly those that are safety-critical. Existing systems (e.g. air traffic control) have some Cloud-like features (load balancing, hot failover, elasticity) but implemented in a specific way, not generally. An advantage of a Cloud environment is that – given appropriate standards – the complete service could be transferred from one Cloud environment to another so ensuring business continuity.

REFERENCES & SOURCES

[1] Malis, A. (1993), 'Routing over Large Clouds (ROLC) Charter", part of the 32nd IETF meeting minutes' - available at *http://www.ietf.org/proceedings/32/charters/rolc-charter.html*

[2] New York Times (2001), 'Internet Critic Takes on Microsoft' - available at *http://www.nytimes.com/2001/04/09/technology/09HAIL.html?ex=1217563200&en=7c46bdefb6a8450a&ei=5070*

[3] Wikipedia, 'John McCarthy (computer scientist)' - available at http://en.wikipedia.org/wiki/John_McCarthy_ (computer_scientist)

[4] Barr, J. (2006), 'Amazon EC2 Beta' - available at http://aws.typepad.com/aws/2006/08/amazon_ec2_beta.html

[5] Sutter, H. (2005), 'The Free Lunch Is Over: A Fundamental Turn Toward Concurrency in Software', in *Dr. Dobb's Journal*, **30**(3).

[6] Toffler, A. (1980), 'The Third Wave', *Pan Books*

[7] Wikipedia, 'Cloud Computing' - available at *http://en.wikipedia.org/wiki/Cloud_computing*

[8] Wikinomics, 'The Prosumers'- available at http://www.socialtext.net/wikinomics/index.cgi?the_prosumers

[9] Golden; B. (2009), 'Capex vs. Opex: Most People Miss the Point About Cloud Economics' - available at *http://www.cio.com/article/484429/Capex_vs._Opex_Most_People_Miss_the_Point_About_Cloud_Economics*

[10] Fellows, W. (2009), 'The State of Play: Grid, Utility, Cloud' - available at http://old.ogfeurope.eu/uploads/Industry%20Expert%20Group/FELLOWS_CloudscapeJan09-WF.pdf

[11] Sims, K. (2009), 'IBM Blue Cloud Initiative Advances Enterprise Cloud Computing' - available at *http://www-03.ibm.com/press/us/en/pressrelease/26642.wss*

[12] Zimory GmbH (2009), 'Zimory Enterprise Cloud – Whitepaper' - available at http://www.zimory.com/fileadmin/images/content_images/pdf/WP_Enterprise_Engl_020409.pdf

[13] RightScale Inc. (2009), 'RightScale Cloud Management Features' - available at *http://www.rightscale. com/products/features/*

[14] DeCandia, G.; Hastorun, D.; Jampani, M.; Kakulapati, G.; Lakshman, A.; Pilchin, A.; Sivasubramanian, S.; Vosshall, P. & Vogels, W. (2007), 'Dynamo: Amazon's Highly Available Key-value Store' - available at *http://s3.amazonaws.com/ AllThingsDistributed/sosp/amazon-dynamo-sosp2007.pdf*

[15] Amrhein, D. & Willenborg, R. (2009), 'Cloud computing for the enterprise, Part 3: Using WebSphere CloudBurst to create private clouds' - available at *http://www.ibm.com/developerworks/websphere/techjournal/0906_amrhein/0906_amrhein.html*

[16] Chappell, D. (2008), 'Introducing the Azure Services Platform' - available at *http://download.microsoft. com/download/e/4/3/e43bb484-3b52-4fa8-a9f9-ec60a32954bc/Azure_Services_Platform.pdf*

[17] Foster, I. (2008), 'Cloud, Grid, what's in a name?' - available at *http://ianfoster.typepad.com/blog/2008/08/cloud-grid-what.html*

[18] Members of EGEE-II (2008), 'An egee comparative study: Grids and clouds - evolution or revolution. Technical report, Enabling Grids for E-sciencE Project' - available at *https://edms.cern.ch/document/ 925013/*.

[19] Vaquero, L. M.; Rodero-Merino, L.; Caceres, J. & Lindner, M. (2009), 'A break in the clouds: towards a cloud definition', *SIGCOMM Comput. Commun. Rev.* **39**(1), 50--55.

[20] Harris, D. (2008), 'Grid vs. Cloud vs. What Really Matters' - available at http://www.on-demandenterprise.com/blogs/Grid-vs-Cloud-vs-What-Really-Matters.html

[21] European Commission, Renewable Energies Unit (2008), 'Code of Conduct on Data Centres Energy Efficiency, Version 1.0' – available at *http://re.jrc.ec.europa.eu/energyefficiency/pdf/CoC%20data%20centres%20nov2008/CoC%20DC%20v%201.0%20FINAL.pdf*

[22] Barroso, L. A.; Hoelzle, U (2009), "The Datacenter as a Computer", Morgan and Claypool Publishers

[23] "Open Cloud Manifesto" - available at *http://www.opencloudmanifesto.org/Open%20Cloud%20Manifesto.pdf*

[24] Vambenepe, W (2009), "Reality check on Cloud portability" - available at *http://stage.vambenepe.com/archives/684*

[25] Petry, A (2007), "Design and Implementation of a Xen-Based Execution Environment" - available at *http://www.xenbee.net/_media/thesis.pdf?id=XenBEE&cache=cache*

[26] Siegler, MG (2009), 'Animoto Is Already Cash-Flow Positive, Raises Another Round To Go To 11' - available at http://www.techcrunch.com/2009/06/17/animoto-is-already-cash-flow-positive-but-raises-another-round-to-go-to-11/

[27] Chong, F; Carraro, G & Wolter, R (2006), 'Multi-Tenant Data Architecture' - available at *http://msdn.microsoft.com/en-us/library/aa479086.aspx*

[28] Leung, AW; Pasupathy, S; Goodson, G & Miller, EL (2008), 'Measurement and analysis of large-scale network file system workloads', *ATC'08: USENIX 2008 Annual Technical Conference on Annual Technical Conference*, USENIX Association, p. 213-226

[29] Next Generation GRIDs Expert Group (2006), 'Future for European Grids: GRIDs and Service Oriented Knowledge Utilities - Next Generation GRIDs Expert Group Report 3', available at *ftp://ftp.cordis.lu/pub/ist/docs/grids/ngg3_eg_final.pdf*

[30] Fan, X; Weber, WD & Barroso, LA (2007), 'Power Provisioning for a Warehouse-sized Computer'. Proceedings of the 34th International Symposium on Computer Architecture in San Diego, CA. Association for Computing Machinery, ISCA '07 – available at *http://labs.google.com/papers/power_provisioning.pdf.*

[31] Clarke, G (2005), 'Open source taking over Europe - We just don't know it' - available at *http://www.theregister.co.uk/2005/10/21/opensource_government/*

[32] Truffle Capital (2007), 'Truffle Capital: European Commission Recognises the Need for a "European Strategy for Software" - Commenting on the 2007 Truffle 100 Europe, Viviane Reding Calls on Europe to Develop a Leadership Position in Software' - available at *http://www.businesswire.com/news/google/20071122005070/en*

[33] DG Information Society and Media - Directorate for Converged Networks and Service (2009), 'Towards A European Software Strategy - Report Of An Industry Expert Group' - available at *http://www.nessi-europe.com/Nessi/LinkClick.aspx?fileticket=7teEO5hzywY%3D&tabid=304&mid=1571*

[34] Webhosting Unleashed (2008), 'Cloud-Computing Services Comparison Guide' - available at *http://www.webhostingunleashed.com/whitepaper/cloud-computing-comparison/*

[35] Wayner, P (2008), 'Cloud versus cloud: A guided tour of Amazon, Google, AppNexus, and GoGrid' - available at *http://www.infoworld.com/d/cloud-computing/cloud-versus-cloud-guided-tour-amazon-google-appnexus-and-gogrid-122?page=0,0*

[36] Golden, B (2009), 'The Cloud as Innovation Platform: Early Examples' - available at *http://www.nytimes.com/external/idg/2009/06/18/18idg-the-cloud-as-innovation-platform-early-examples-24294.html*

[37] Schubert, L; Kipp, A; & Wesner, S (2009), 'Above the Clouds: From Grids to Resource Fabrics'. In G. Tselentis, J. Domingue, A. Galis, A. Gavras, D. Hausheer, S. Krco, et al., *Towards the Future Internet - A European Research Perspective* (pp. 238 - 249). Amsterdam: IOS Press.

[38] Mell, P & Grance, T (2009), 'National Institute of Standards and Technology, Information Technology Laboratory', *http://groups.google.com/group/cloudforum/web/nist-working-definition-of-cloud-computing*

[39] Armbrust, M; Fox, A; Griffith, R; Joseph, AD; Katz, RH; Konwinski, A; Lee, G; Patterson, DA; Rabkin, A; Stoica, I & Zaharia, M (2009), 'Above the Clouds: A Berkeley View of Cloud Computing'. Technical Report No. UCB/EECS-2009-28 – available at *http://www.eecs. berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html*

[40] Rochwerger, R; Caceres, J; Montero, RS; Breitgand, D; Elmroth, E; Galis, A; Levy, E; Llorente, IM; Nagin, K & Wolfsthal, Y (2009), 'The RESERVOIR Model and Architecture for Open Federated Cloud Computing'. IBM Systems Journal, September 09

[41] OpenNebula: Sotomayor, B; Montero, RS; Llorente, IM & Foster, I (2009), 'An Open Source Solution for Virtual Infrastructure Management in Private and Hybrid Clouds'. IEEE Internet Computing, Special Issue on Cloud Computing, October 09

[42] Challengers (2009), 'Final Research Agenda on Core and Forward Looking Technologies' – available at *http://challengers-org.eu/index.php/Download-document/57-01.-CHALLENGERS-Research-Agenda-and-Roadmap-Final-Version-January-2009.html*

[43] The NESSI-Grid Project (2008), 'Grid Vision and Strategic Research Agenda' – available at *http://www.soi-nwg.org/lib/exe/fetch.php?id=start&cache=cache&media=nessi_grid_sra_ v3.0.pdf*

[44] Massó, J (2009), 'Stormy Weather (Cloud & SaaS)'. INES General Assembly Keynotes – available at *http://www.ines.org.es/docs/4%20Asamblea/06%20-%20StormyWeather_ jmasso_INES_8Jul.pdf*

[45] See e.g. ycombinator.com, or http://www.pdc.kth.se/Members/edlund/ISSGC09-Aake-Edlund.pdf

[46] bgin.wordpress.com (BG Innovation Lab)

[47] www.necloud.org, a Northern Europe project collaborating with UK, NL, Spain and Greece

[48] Foster, I (1998), 'The Grid: Blueprint for a New Computing Infrastructure', Morgan Kaufmann Publishers

[49] Next Generation GRIDs Expert Group (2003), 'Next Generation GRIDs: European Grid Research 2005-2010', available at *ftp://ftp.cordis.lu/pub/ist/docs/ngg_eg_final.pdf*

[50] Asanovic, K; Bodik, R; Catanzaro, BC; Gebis, JJ; Husbands, P; Keutzer, K; Patterson, DA; Plishker, WL; Shalf, J; Williams, SW; Yelick, KA (2006), 'The Landscape of Parallel Computing Research: A View from Berkeley'. Technical Report No. UCB/EECS-2006-183 – available at *http://www.eecs. berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.pdf*

[51] European Commission (2007), 'building the e-Infrastructure: Computer and network infrastructures for research and education in Europe. A pocket guide to the activities of the Unit GÉANT & e-Infrastructure' – available at *ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/e-infrastructure/ leaflet-2006-building-e-infrastructure_en.pdf*

[52] Brandic, I; Music, D; Leitner, P; Dustdar, S (2009), 'VieSLAF Framework: Enabling Adaptive and Versatile SLA-Management'. Gecon09. In conjunction with Euro-Par 2009, 25- 28 August 2009, Delft, The Netherlands.

[53] Kincaid, J (2009), 'T-Mobile Sidekick Disaster: Danger's Servers Crashed, And They Don't Have A Backup', available at *http://www.techcrunch.com/2009/10/10/t-mobile-sidekick-disaster-microsofts-servers-crashed-and-they-dont-have-a-backup/*

[54] Catteddu, D; Hogben, G eds. (2009), 'Cloud Computing - Benefits, risks and recommendations for information security', European Network and Information Security Agency (ENISA) – available at *http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport*

[55] Barham, P; Dragovic, B; Fraser, K; Hand, S; Harris, T; Ho, A; Neugebauer, R; Pratt, I & Warfield, A (2003), 'Xen and the Art of Virtualization', Technical report, University of Cambridge Computer Laboratory – available at http://www.cl.cam.ac.uk/research/srg/netos/papers/2003-xensosp.pdf

The Future Of Cloud Computing

Opportunities For European Cloud Computing Beyond 2010

••• Expert Group Report

Public Version 1.0

Rapporteur for this Report: Lutz Schubert [USTUTT-HLRS]

Editors: Keith Jeffery [ERCIM], Burkhard Neidecker-Lutz [SAP Research]

European Commission
Information Society and Media