MTA SZTAKI

Cloud computing

Dr. Bakonyi Péter
.

1/24/2011

MTA SZTAKI

I DEFINITIONS

cloud (klaʊd) *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).*

MTA SZTAKI

Cloud definició

- A cloud vagy felhő egy platform vagy infrastruktúra
- Az alkalmazások és szolgáltatások végrehajtására alkalmas
- Az előre definiált minőségi paraméterek automatikusan teljesülnek
- Az erőforrások skálázhatók és az aktuális kivánalmaknak felelnek meg

MTA SZTAKI

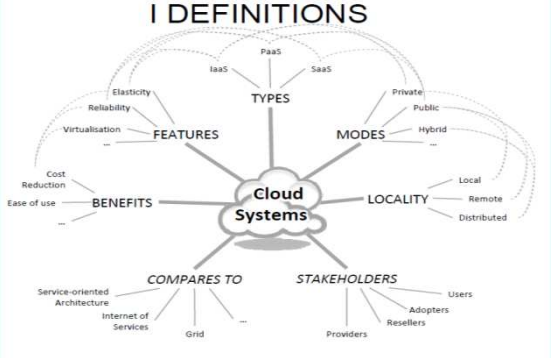
Cloud definició

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.

MTA SZTAKI

I DEFINITIONS



MTA SZTAKI

- Stakeholders (érintett résztvevők):
  Szolgáltatók, felhasználók , értékesítők
- Előnyök:
  költség csökkentés
  egyszerű használat

MTA SZTAKI

Virtualizáció

•A virtualizáció jelenetése hogy egy estzköznek vagy erőforrásnak virtuális verzióját állitom elő.

Pl. szerver, tároló, operációs rendszer. A keretprogramaz erőforrásokat egy vagy több végrehajtási környezetre képezi le.

Az ez emberek, eszközök, az alkalmazások interakcióba lépnek a virtuális erőforrásokkal mint valós logikai egységekkel.

MTA SZTAKI

Tuljdonságok:

Virtualization

•In computing, virtualization means to create a virtual version of a device or resource, such as a server, storage device, network or even an operating system where the framework divides the resource into one or more execution environments. Even something as simple as partitioning a hard drive is considered virtualization because you take one drive and partition it to create two separate hard drives. Devices, applications and human users are able to interact with the virtual resource as if it were a real single logical resource. The term virtualization has become somewhat of a buzzword, and as a result the term is now associated with a number of computing technologies including the following:

•storage virtualization: the amalgamation of multiple network storage devices into what appears to be a single storage unit.

•server virtualization: the partitioning a physical server into smaller virtual servers.

•operating system-level virtualization: a type of server virtualization technology which works at the operating system (kernel) layer.

•network virtualization: using network resources through a logical segmentation of a single physical network.

•application virtualization

MTA SZTAKI

•Rugalmasság

A szolgáltató szempontjából:

A rendszer kapacitások automatikus igény szerinti növelése vagy csökkentése - CPU, memória , I/O , sávszélesség-

A szolgáltató határozza meg az automatizálás fokát. Kézi hozzáadás nem támogatott, de megtörténik

•Valós idejű infrastruktúra jelentése osztott IT infrastruktúra-ahol az IT erőforrások automatikus és optimalizált hozzárendelését a üzleti stratégis és az SLA határozza meg

MTA SZTAKI

Elasticity

Analysis By: Daryl Plummer

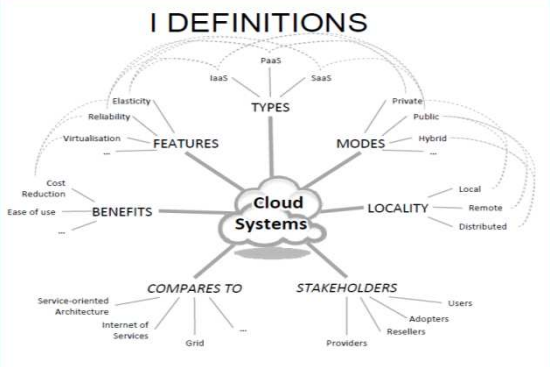
Definition: In the service provider view, cloud service elasticity is the ability to increase or decrease the amount of system capacity (for example, CPU, storage, memory and input/output [I/O] bandwidth) that is available to a given cloud service on demand in an automated fashion. The degree of automation of elasticity is determined by the service provider. Manual means of adding capacity are generally discouraged, but are more commonplace in 2009.

Real-Time Infrastructure

Analysis By: Donna Scott

Definition: Real-time infrastructure (RTI) represents a shared IT infrastructure (across customers, business units or applications) in which business policies and service-level agreements drive dynamic and automatic allocation and optimization of IT resources (that is, services are elastic), so that service levels are predictable and consistent, despite the unpredictable demand for IT services. Where resources are constrained, business policies determine how resources are allocated to meet business goals. RTI may be implemented in private and public cloud architectures, and where it is implemented is what provides the elasticity functionality.

MTA SZTAKI



MTA SZTAKI

A cloud tipusai

•Infrastruktúra mint szolgáltatás

•Erőforrás cloud- megnövelt virtualizációs képesség. Az erőforrásokat egy szolgáltatói interfacen keresztül biztosítja

•Platform mint szolgáltatás

•Egy platfolmon nyújt számítástechnikai erőforrást, alkalmazás és szolgáltatás fejlesztés céljára. Tipikusan ezt egy dedikált API-n keresztül biztosítja. Ez változik a szolgáltatóktól függöen

MTA SZTAKI

•Szoftver mint szolgáltatás - SaaS

Úgy is nevezik,hogy Szolgáltatás vagy Alkalmazás Cloud

Biztosítja az alkalmazások és szolgálztatásokhoz szükséges cloud infrastruktúrát, vagy platformot.

Gyakran szabványos alaklmazási szoftvereket ajánlanak cloud szolgáltatásként

MTA SZTAKI

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]).

(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:

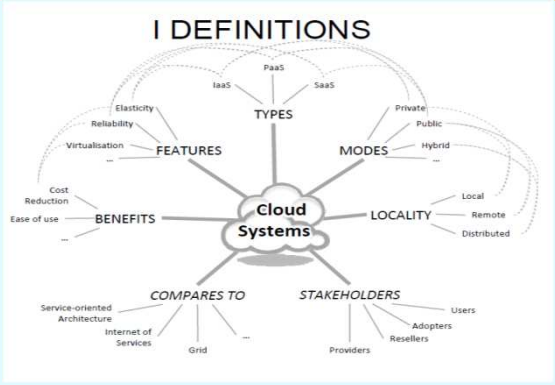
(Cloud) Platform as a Service (PaaS), provide computational resources via a *platform* upon which applications and services can be developed and hosted. PaaS typically makes use of dedicated APIs to control the behaviour of a server hosting engine which executes and replicates the execution according 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).

MTA SZTAKI

(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.

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.

MTA SZTAKI



MTA SZTAKI

Modes-Cloud usage

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.

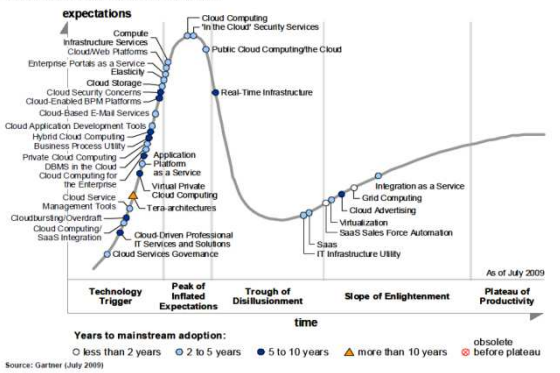
Public Clouds. Enterprises may use cloud functionality from others, respectively offer their own services to users outside of the company. Providing the user with the actual capability to exploit the cloud features for his / her own purposes also allows other enterprises to outsource their services to 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.

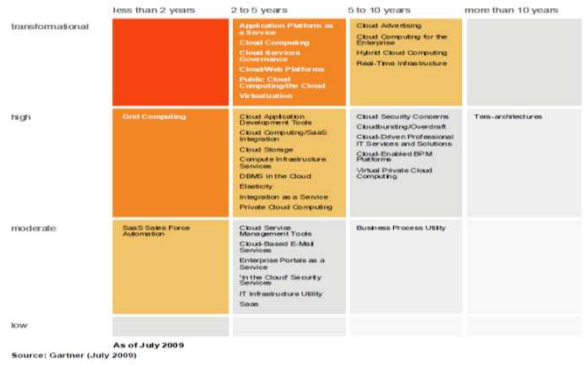
MTA SZTAKI

Figure 1. Hype Cycle for Cloud Computing, 2009



MTA SZTAKI

Figure 2. Priority Matrix for Cloud Computing, 2009



MTA SZTAKI

Cloud Services Governance

Analysis By: Frank Kenney; Daryl Plummer; Benoit Lheureux

Definition: Cloud services governance (CSG) is a discipline focused on coordinating, monitoring and management of cloud service interactions. These interactions may take place from service to service, from service provider to service consumer, or across collections of services regardless of whether the governance technologies are deployed as a service offering or on-premises. The discipline involves six characteristics of governance that are significant to cloud services: visibility, management, monitoring, measurement, provisioning and validation. Within these characteristics,

Cloud-Driven Professional IT Services and Solutions

Analysis By: Frances Karamouzis

Definition: There are two distinctly different areas of cloud-driven professional IT services and solutions.

The first area includes all types of consulting, advisory, deployment and testing services provided to enterprises from professional IT service companies. IT service companies assist clients in understanding and navigating the various areas of cloud computing. This includes business advisory services to strategically help clients determine the potential impact on their business model, options for shifts in their technology architecture or future opportunities. The IT consulting,

MTA SZTAKI

The second area of cloud-driven professional IT services and solutions includes all types of solutions that are developed, bundled, and packaged as outsourcing offerings, where the IT service provider leverages one or more cloud computing technologies within the solution's overall architecture. Gartner refers to these IT services as "cloud-leveraged outsourcing offerings" — for example, a platform business process outsourcing (BPO) offering, where the business process and application layers of the solution are indigenous to the IT service provider, and the infrastructure is delivered through the cloud. With so many elements to each outsourcing layer (BPO, application outsourcing and IT infrastructure), the combination of the different types of cloud-driven professional IT services and solutions is extensive.

Cloudbursting/Overdraft

Analysis By: Daryl Plummer

Definition: One of the key value propositions of cloud computing is the ability to increase or decrease service capacity on demand and to pay for only what you use. This is commonly referred to as "cloud service elasticity." Along with that idea is a complementary idea called "capacity overdrafting" or "cloudbursting" (we use the terms interchangeably). It is the ability to automatically get more capacity from a different cloud infrastructure when the primary cloud infrastructure is overloaded (see "Anatomy of a Cloud 'Service Overdraft': One Way Elasticity Happens").

MTA SZTAKI

Cloud Service Management Tools

Analysis By: Cameron Haight; Milind Govekar

Definition: Cloud service management tools are products that provide visibility and control within external (public) cloud environments to enterprise consumers and IT operations teams. Included in this category are products that provision system images, monitor performance and availability, enable metering and billing, and integrate with enterprise management systems.

Virtual Private Cloud Computing

Analysis By: Lydia Leong; David Cearley

Definition: A virtual private cloud (VPC) refers to the partitioning of a portion of a public cloud computing service provider's environment into an isolated environment that is dedicated for use by a single entity or group of related entities (such as multiple departments within a company). In addition, a VPC may be isolated from the Internet, utilizing a private network (virtual private network [VPN] or private connectivity) and/or a virtual LAN for access to the services, to add an additional level of performance, security and control.

Application Platform as a Service

Analysis By: Yefim Natis; Eric Knipp

Definition: Application platform as a service (APaaS) is a development and deployment environment for cloud-based applications, offered to IT organizations as a service. In other words, applications developed in and for an APaaS are software-as-a-service (SaaS) applications.

APaaS is, in principle, a specialized application server and application development toolset that is deployed "in the cloud" and offered as a service to its users. The technology behind the service is referred to as a SaaS-enabled application platform (SEAP). Some SEAPs are offered as general-purpose products, while others are used exclusively to power an APaaS and are not offered as products in their own right. The cloud specialty of such an application server makes it quite different from a regular application server. In addition to the underlying SEAP technology used to enable APaaS, some vendors also provide highly scalable, distributed data stores and other innovations that differ substantially from traditional on-premises alternatives.

DBMS in the Cloud

Analysis By: Donald Feinberg

Definition: Database management systems (DBMSs) in the cloud consist of any DBMS available on a cloud infrastructure. There are three distinct variations of DBMSs in the cloud:

MTA SZTAKI

Private Cloud Computing

Analysis By: Thomas Bittman

Definition: The term "private cloud computing" describes a style of computing used by a modern internal IT provider to behave like an external, cloud-computing service provider (see "Private Cloud Computing: The Steppingstone to the Cloud"). Private cloud computing is a style of computing in which scalable and elastic IT-enabled capabilities are delivered as services to "internal" customers using Internet technologies.

Private cloud computing is the answer for larger enterprises interested in improving their economies of scale and efficiency, improving their flexibility and elasticity, and lowering their barrier to entry to IT customers, without necessarily relying on "public" cloud-computing services (because the services are not yet available, mature or secure, for example).

MTA SZTAKI

Cloud Security Concerns

Analysis By: Jay Heiser; Arabella Hallawell; David Cearley

Definition: Today's cloud computing services are relatively nontransparent, making it extremely difficult for potential customers to assess the relative security and compliance risks. Best practices for the risk assessment and security control of cloud offerings have yet to be established or require nascent third-party security controls, leaving most organizations ignoring security requirements, avoiding cloud computing entirely or experimenting with unproven techniques.

Elasticity

Analysis By: Daryl Plummer

Definition: In the service provider view, cloud service elasticity is the ability to increase or decrease the amount of system capacity (for example, CPU, storage, memory and input/output [I/O] bandwidth) that is available to a given cloud service on demand in an automated fashion. The degree of automation of elasticity is determined by the service provider. Manual means of adding capacity are generally discouraged, but are more commonplace in 2009.

MTA SZTAKI

Compute Infrastructure Services

Analysis By: Lydia Leong

Definition: Compute infrastructure services offer on-demand computing capacity from a service provider. Rather than buying servers and running them within its own data center, a business simply obtains the necessary infrastructure from a service provider in a shared, scalable, "elastic" way and accesses it via the public Internet or a private network.

'In the Cloud' Security Services

Analysis By: Kelly Kavanagh; Greg Young

Definition: "In the cloud" security services are Internet-fabric-based managed firewalls, intrusion detection systems, intrusion prevention systems, antivirus services, distributed denial-of-service protection services, messaging security and Web gateway security services.

MTA SZTAKI

Cloud Computing

Analysis By: David Mitchell Smith

Definition: Gartner defines "cloud computing" as a style of computing where scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.

Position and Adoption Speed Justification: Users are changing their buying behaviors. Although it is unlikely that they will completely abandon on-premises models, or that they will soon buy complex, mission-critical processes as services through the cloud, there will be a movement toward consuming services in a more cost-effective way. As expected of something at the Peak of Inflated Expectations, there is deafening hype around cloud computing. Every IT vendor has a cloud strategy, although many aren't cloud-centric. Variations, such as private cloud computing and hybrid approaches, compound the hype and demonstrate that one dot on a Hype Cycle cannot adequately represent all that is cloud computing.

MTA SZTAKI

Public Cloud Computing/the Cloud

Analysis By: Daryl Plummer

Definition: Gartner's definition of cloud computing essentially describes public cloud computing as a style of computing where scalable and elastic IT-enabled capabilities are provided "as a service" to external customers using Internet technologies. Therefore, public cloud computing is the use of cloud-computing technologies to support customers that are external to the organization of the provider. It is through public consumption of cloud services that the types of economies of scale and the sharing of resources will be generated to reduce cost and to increase choices available to consumers.

Real-Time Infrastructure

Analysis By: Donna Scott

Definition: Real-time infrastructure (RTI) represents a shared IT infrastructure (across customers, business units or applications) in which business policies and service-level agreements drive dynamic and automatic allocation and optimization of IT resources (that is, services are elastic), so that service levels are predictable and consistent, despite the unpredictable demand for IT services. Where resources are constrained, business policies determine how resources are allocated to meet business goals. RTI may be implemented in private and public cloud architectures, and where it is implemented is what provides the elasticity functionality.

MTA SZTAKI

SaaS

Analysis By: Robert DeSisto

Definition: Software as a service (SaaS) is software that is owned, delivered and managed remotely by one or more providers. If the vendor requires user organizations to install software on-premises using their infrastructures, then the application isn't SaaS. SaaS delivery requires a vendor to provide remote, outsourced access to the application, as well as maintenance and upgrade services for it. The infrastructure and IT operations supporting the applications must also be outsourced to the vendor or another provider.

Virtualization

Analysis By: Thomas Bittman

Definition: IT virtualization is the abstraction of IT resources in a way that masks the physical nature and boundaries of those resources from resource users. An IT resource can be a server, a client, storage, networks, applications, operating systems or a search engine. Essentially, any IT building block can potentially be abstracted from resource users.

MTA SZTAKI

Cloud Advertising

Analysis By: Andrew Frank

Definition: Cloud advertising is a business process cloud service defined as the capability to deliver advertising where the content and the fee charged is determined at the time of end-user access, usually by an auction mechanism that matches bidders with "spots" as they become available. Search engine marketing (SEM) and various forms of online display advertising (e.g., banners) are the most-developed formats, but the concept is also evolving to other channels and platforms such as online video, mobile devices, addressable television, and out-of-home digital signage.

Grid Computing

Analysis By: Carl Claunch; Andrew Butler

Definition: Grid computing refers to using computers managed by more than one organization, whether internal or external, to collectively accomplish large tasks, such as derivative risk analysis, candidate drug screening or complex simulations.

MTA SZTAKI

Multitenancy

•**Multitenancy** refers to a principle in software architecture where a single instance of the software runs on a server, serving multiple client organizations (tenants). Multitenancy is contrasted with a multi-instance architecture where separate software instances (or hardware systems) are set up for different client organizations. With a multitenant architecture, a software application is designed to virtually partition its data and configuration so that each client organization works with a customized virtual application instance.

MTA SZTAKI

Data management

•The official definition provided by DAMA or Boucher: "Data Resource Management is the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise." This definition is fairly broad and encompasses a number of professions which may not have direct technical contact with lower-level aspects of data management, such as relational database management.

•Alternatively, the definition provided in the DAMA Data Management Body of Knowledge (DAMA-DMBOK) is: "Data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.

MTA SZTAKI

Hybrid Cloud Computing

Analysis By: David Cearley

Definition: For the near future, virtually all companies using public cloud-computing services will also have some form of internal IT systems. However, hybrid cloud computing does not refer to using internal systems and external cloud-based services in a disconnected or loosely connected fashion. Hybrid cloud computing refers to the combination of external public cloud-computing services and internal resources (either a private cloud or traditional infrastructure, operations and applications) in a coordinated fashion to assemble a particular solution. Hybrid cloud computing implies significant integration or coordination between the internal and external environments at the data, process, management or security layers.

Hybrid cloud computing can take a number of forms. These approaches can be used individually or in combination to support a hybrid cloud-computing approach:

- **Joint security and management** — Security and/or management processes and tools are applied to the creation and operation of both internal systems and external cloud services.
- **Cloudbursting** — Dynamically extending an application or a portion thereof from an internal private cloud platform to an external public cloud service based on the need for additional resources.
- **Cloud service composition** — Creating a solution with a portion running on internal systems, and another portion delivered from the external cloud environment in which there is ongoing data exchanges and process coordination between the internal and external environments. Mashups are a form of integrated solutions where public cloud-

MTA SZTAKI

Capabilities of Cloud

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.

MTA SZTAKI

Capabilities of Cloud

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.

MTA SZTAKI

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)

MTA SZTAKI

2. ECONOMIC ASPECTS OVERVIEW

	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 "Going Green"	general issue		No dedicated tool support			
	☑ 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)

MTA SZTAKI

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 (engine) ☐ 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)

MTA SZTAKI

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.

MTA SZTAKI

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.

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.

MTA SZTAKI

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

MTA SZTAKI

Köszönöm a figyelmet!