ETSI TR 102 997 V1.1.1 (2010-04)

Technical Report

CLOUD;
Initial analysis of standardization requirements for Cloud services

Reference

DTR/GRID-0009 StdRqmtsCloudSvc

Keywords

service, interoperability, ICT, management

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Copyright Notification

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee CLOUD (CLOUD), formerly TC GRID (GRID).

1 Scope

The present document describes standardisation requirements for cloud services. It is based on the outcome of the ETSI TC GRID Workshop, "Grids, Clouds and Service Infrastructures", 2 and 3 December 2009. This event brought together key stakeholders of the grid, cloud and telecommunication domains to review state of the art and current trends. Needs for standardisation, with a particular focus on the emerging area of cloud computing and services, were discussed. The present document introduces and expands on the conclusions reached. This is not an exhaustive survey and is intended to serve as the basis for future work.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

Not applicable.

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area. For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] The NIST Definition of Cloud Computing, Peter Mell and Tim Grance, Version 15, 10 July 2009.

NOTE: See http://csrc.nist.gov/groups/SNS/cloud-computing/.

[i.2] Workshop on "Grids, Clouds and Service Infrastructures", 1-3 December 2009, ETSI, Sophia Antipolis, France.

NOTE: See http://www.etsi.org/plugtests/GRID09/GRID.htm.

[i.3] Open Virtualization Format Specification, DMTF Document Number DSP0243, 12 January 2010.

NOTE: See http://www.dmtf.org/initiatives/vman_initiative.

[i.4] "Cloud Computing and the Internet", Vinton Cerf, Google Research Blog, 28 April 2009.

NOTE: See http://googleresearch.blogspot.com/2009/04/cloud-computing-and-internet.html.

[i.5] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

[i.6] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing or personal data and on the free movement of such data.

3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

API	Application Programming Interface
DMTF	Distributed Management Task Force
ICT	Information and communication technologies
LI	Lawful Interception
NIST	National Institute of Standards and Technology
OVF	Open Virtualisation Format
PaaS	Platform-as-a-Service
SLAs	Service Level Agreements
VPN	Virtual Private Networks
DPP	Data Protection and Privacy

4 Cloud computing - an introduction

The term "cloud computing" is applied to a range of different approaches to delivering IT capabilities over networks, typically the Internet. It originates from the use of a cloud to represent wide area networks in diagrams - indicating that the details of how data are transported are hidden from the endpoints. Only correct delivery is significant to the end user. The network provider has the freedom to configure its systems and operations to meet its own business goals. Cloud computing represents the extension of this general idea to include a wider range of networked IT components such as servers, storage and data resources.

The National Institute of Standards and Technology (NIST) has recently proposed a definition of cloud computing which is becoming generally accepted:

- "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [i.1].

Various services based on cloud computing infrastructures have emerged recently. From the customers' point of view, much of their attraction lies in the ability to purchase only what they need, infrastructure, value-added platforms or software packages, without having to plan far ahead. There is typically little up-front commitment and the potential ability to achieve flexible scaling to meet dynamic demand. This can enable "pay-as-you-grow" commercial models.

The shift to utility models such as cloud computing can be compared with the move from private circuits to Virtual Private Networks (VPN). Moving from private, dedicated infrastructure to a managed shared service promises clear cost benefits to the customer, provided that issues such as performance, availability and security are predictable and meet their needs. From a provider's point of view, sharing of resources between customers and bulk purchasing can lead to economies of scale. Commercial trends in computing infrastructure have reached a point where these economies of scale can make provision of computing infrastructure as a utility viable. However, mainstream adoption of cloud services is currently limited, at least in part due to the technological diversity of current offerings in terms of e.g. different virtualization technologies or different interface definitions (API's) towards the services. The following clauses describe requirements for standards to promote the development of the market for cloud services. This is based on the conclusions of the ETSI TC GRID Workshop [i.2].

5 Standards Requirements for Cloud Computing

5.1 Portability

Portability in general refers to the ability to migrate applications between different clouds. This is required to allow customers of cloud services to avoid the situation of being locked into a specific cloud infrastructure provider, having made the decision to run an application in the cloud. This adds to the perceived risks associated with moving to cloud computing. A potential customer needs a high level of trust in the technical and commercial ability of a chosen provider to support critical business applications in the long term.

Current cloud infrastructure providers offer their own proprietary interfaces to application developers. Standardised interfaces to manage cloud infrastructures and the different types of resource they provide are required. Reducing the mismatch between different cloud infrastructure systems would not only enable a competitive market but also enable new business models where different cloud infrastructures can be traded according to price and demand.

At one extreme, the ability to automatically migrate a complete running application (including any necessary monitoring and management features) from one cloud infrastructure to another would clearly be attractive to customers (but much less so to cloud infrastructure providers, particularly in the current market).

Portability of a virtual machine images is being addressed by the DMTF Open Virtualisation Format (OVF) [i.3]. This should provide a good basis for limited portability but does not address complex configuration or interactions with any supporting systems. These issues are naturally within the scope of a service provider.

Portability of data is essential for a typical business application. The ability for a customer to retrieve application data from one cloud infrastructure provider and import this into an equivalent application hosted by an alternative provider reduces the risk of long-term dependency. This would probably involve a not insignificant amount of effort from the customer (or an application service provider - the details will be largely application specific and not the responsibility of a cloud infrastructure provider). Achieving data portability depends on effective standardisation of data import and export functionality between cloud infrastructure providers.

5.2 Interoperability of clouds

Interoperability is closely related to portability. Here we interpret interoperability as the ability to federate multiple clouds to support a single application. In other words, interoperability involves software and data simultaneously active in more than one cloud infrastructure, interacting to serve a common purpose.

Considering cloud interoperability in general, the term "Intercloud", possibly coined by Cisco ^{TM}, is starting to gain some acceptance. It is analogous to the Internet and based on a similar vision - connecting individual, essentially uncoordinated cloud infrastructures and giving control to the users. One of the lessons of the Internet is that the use of common protocols (or interfaces) readily accessible to developers has great benefits in stimulating innovation. In a recent lecture, Vinton Cerf discussed the current state of play [i.4]:

- "Each cloud is a system unto itself. There is no way to express the idea of exchanging information between distinct computing clouds because there is no way to express the idea of ‘another cloud'. Nor is there any way to describe the information that is to be exchanged. Moreover, if the information contained in one computing cloud is protected from access by any but authorized users, there is no way to express how that protection is provided and how information about it should be propagated to another cloud when the data is transferred".

There are two obvious scenarios that have to be considered when an application is supported by a federation across multiple clouds. In the first scenario the application is managed by an entity, which may be the end user, which interacts individually with each cloud provider. This entity is directly responsible for coordination between the cloud service providers and manages the application via standardised management and monitoring interfaces. In the second scenario a cloud service provider may take on this coordination role, using services from a number of third party providers to meet the overall performance guarantee to the end user. One example could be a compute cloud provider using a network provider for delivery. This scenario requires standards for the sharing of management information such as SLA goals between service providers without exposing too much confidential detail of how the goals are met.

Even these two quite obvious scenarios allow us to suggest a number of dimensions for the discussion of Cloud interoperability:

- **Application/Service:** Interoperability standards should support distributed applications with predictable behaviour and performance. Components of a single application could be deployed across multiple cloud infrastructure providers and possibly reconfigured while running, or with limited interruption, to respond to changes in usage patterns or resource availability, for example. Application configuration should be resilient to changes in the configuration within each cloud - for example scaling or migration of computational resources.

- **Management:** Standardised interfaces should be provided by cloud service providers so that a single application can be managed in a consistent way, end-to-end - substantially independent of the details of its deployment across multiple cloud infrastructures. It will be possible for a management application to control and coordinate components in multiple clouds Standardised management functionality for deployment and migration of virtual machine images between different cloud infrastructures is required. Management interoperability requires interactions between multiple independent actors responsible for application management and infrastructure management.

- **Data:** Standards are required to support the deployment of equivalent virtual machine images and application data to different cloud infrastructures. A basic requirement for cloud interoperability is network connectivity between cloud environments, appropriate to carry application traffic. In particular, security and other cross-domain data management issues need to be handled in standard ways.

- **Network aspects:** Network access to computational resources is fundamental to cloud services. Standards are required to support both uniform access to individual cloud computing resources and concatenation or federation of clouds in different locations. Interconnection between clouds should support the quality requirements of applications. Network connectivity will in general be based on the use of shared physical resources in a similar way to computing and storage. Standards for allocation and admission control will be needed.

5.3 Closer integration of IT and network resources

The anticipated convergence of computing, storage and networking into a single integrated infrastructure is becoming a reality. The distinction between network services and the applications they support is disappearing. From the user's point of view, the application experience is what matters and this depends on all the supporting systems performing effectively. This closer integration of IT and network resources is of special interest for real-time and interactive applications (e.g. from the telecommunications area) with particular requirements on network performance. Deploying these applications in the cloud will become feasible when IT and network resources are integrated as a unified infrastructure.

The trend towards cloud computing is breaking the association between application software and physical hardware in the interests of flexibility and resource efficiency - a shift already underway as a result of server virtualisation. As hardware configurations in complex applications change dynamically, the task of keeping all the endpoints connected becomes increasingly complicated.

A consistent approach to automation of ICT infrastructure is required. Each component (computing, storage and network) needs to be dynamically configurable to respond to changes in the performance or configuration of other elements. Effective monitoring and control solutions are needed, which allow the automation and associated cost benefits offered by cloud computing infrastructures to be extended across the whole infrastructure involved in supporting applications.

5.4 APIs to networking/data movement functionality

This requirement is closely related to topics covered in clauses 5.2, 5.3 and 5.7. Today's (value-added/platform) cloud service providers offer relatively static resource management, typically from a single infrastructure provider. Resources such as computing, storage and network are available to their customers (e.g. application providers) from several different providers and they will increasingly expect to be able to make use of the full range of available infrastructure in a dynamic way. This requires the ability for a distributed application to configure or adapt to the resources available to it at runtime. Current efforts to allow applications to interact with the (virtual) computing resources available to them include, for example, scaling by changing the amount of computing resource allocated to an image, or by adding additional virtual machines. Such mechanisms need to be supplemented with functionality to allow efficient and adaptive movement of data over the networks supporting a distributed application. This could be facilitated by providing suitable APIs. They should offer the ability to discover available network connectivity and storage resources, and for applications to use them effectively.

5.5 Support for building, modelling, testing and deploying applications

To promote the wider acceptance of cloud platforms by application providers, it is necessary to make it easier to use those platforms to achieve dependable applications. There should be tool support for application providers giving them the ability to efficiently develop, deploy and manage their applications. It is not clear to what extent such tools will require standardisation (e.g. APIs), but the ability to target multiple platforms from a common toolset may be attractive. Examples of service engineering and support functionality which would be useful in cloud platforms include:

- Performance estimation for components of loosely coupled applications
- Benchmarks to characterise and verify performance parameters of application components
- Monitoring of application performance and relation to infrastructure level quality parameters
- Mapping of high level application descriptions to resource requirements

The aim is to help an application developer and provider to understand how to construct applications in a modular way and deploy onto a cloud infrastructure, to characterise individual components, to model the expected behaviour and validate the results of modelling in controlled test environments prior to live deployment. Once an application is deployed, appropriate monitoring is required to verify that the application behaves as anticipated.

5.6 Support for optimisation of distributed applications

The typical expectations of a customer of cloud services are that their applications deliver well-defined quality levels to end users, independent of the load on the application (e.g. number of active users) and at a price proportional to the load. Application software and data is deployed into one or more cloud infrastructures according to anticipated usage patterns, taking into account expected availability of resources (computing, storage and network) which are shared with other applications. It may be necessary to adjust the deployment of an application if usage patterns or resource availability change.

It is desirable that the stakeholders (e.g. cloud service providers, application provider) responsible for this deployment can make informed decisions to achieve good resource utilisation and assure application quality. This requires knowledge of characteristics of both the application and the cloud infrastructure. Standardisation of the description of these characteristics could play an important role, particularly where an application is not restricted to a single cloud (e.g. hybrid private/public deployments, cloud portability or interoperability scenarios).

5.7 Clearly defined SLAs, fit for business use

Service Level Agreements (SLAs) in this context are understood to be unambiguous statements of the expectations and responsibilities of both users and providers of cloud services.

An SLA is a contract between the provider and the customer of a service specifying the function performed by the service, the agreed bounds of performance, the obligations on both parties to the contract and how any deviations are to be handled. An SLA is made in a business context and therefore will include all aspects of the interaction between the provider and customer relevant to the service.

There are two distinct requirements for standardisation relating to SLAs for cloud services.

- SLAs (or SLA templates, representing an invitation to treat) need representations which are clearly understood by both parties, make explicit the expectations and obligations on each party and can be used to compare different providers.
- A means of mapping between higher level and lower level requirements in a clearly defined way would also be desirable. This could support some form of end-to-end negotiation in cases where a service involves multiple stakeholders and independent bipartite interactions prove insufficient to assure end-to-end service behaviour.

In order to match the technical characteristics of cloud infrastructures with an appropriate commercial environment, a flexible, lightweight and dynamic framework for management throughout the SLA lifecycle is required.

5.8 Data protection, privacy, and security in clouds

This refers to the data protection, privacy and security issues associated with the use of cloud services for sensitive applications. It encompasses processing and storing of personal or otherwise sensitive (such as business sensitive data) data. Processing, using or storing data in clouds is widely perceived as introducing new risks to customers of cloud services. The abstraction associated with the use of cloud infrastructure can result in reduced control by the owner of the data. Widespread adoption of cloud services for enterprise applications will require confidence that potentially sensitive business data is handled appropriately - governed by policies set by the owner of the data. In addition, distribution of responsibilities concerning the processing and storing of personal data is an area that needs to be addressed both on a technical and regulatory level.

Privacy concerns the right of natural persons to not be subject to identity crime, tracking, or other undesired and unlawful intervention concerning an individual's identity or behaviour.

Data protection addresses the aspect of preventing undesired disclosure or manipulation of personal or otherwise sensitive information.

Security covers confidentiality, integrity, availability, authenticity, accountability and non-repudiation.

Cloud infrastructure is based on the principle of sharing and geographical distribution of resources and this enables a new attack venue, as data from different legal entities share the same distributed infrastructure. Sensitive data from several entities may be physically co-located, although logically separated. Sensitive data from a single entity may be spread across several physical locations with different security policies. These are only a few of the many privacy, data protection and security aspects of importance that may be critical for the wider adoption of cloud computing. There are also regulatory aspects putting constraints on cloud adoption, such as the EU directives 2002/58/EC [i.5] "The processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)" and 95/46/EC [i.6] "The protection of individuals with regard to the processing of personal data and on the free movement of such data".

Topics of interest include:

- Protection against communications threats and risks
- Identification, authentication and authorization of users, providers and services
- Service and data protection
- Protection against malicious communication
- Secure storage
- Lawful Interception (LI)

- Data retention

5.9 Regulatory Aspects

Regulation has a broader impact on cloud services than just Data Protection and Privacy (DPP). Some current approaches (e.g. EU Directives) include constraints based on the location and control of data. It is not clear how location is defined in cloud services. Answering this question may require exposing some of the details of how cloud infrastructure is built - in terms of data centre locations and the way that storage and transport of data is managed (placed, distributed, replicated, cached,…), for example. Cloud service providers are likely to be reluctant to share this information openly but mechanisms for demonstrating compliance with relevant legislation, acceptable to regulatory authorities, are required. In addition, accountability for security and privacy of personal data needs to be properly defined - in a cloud deployment, what is the balance of responsibility between the owner of the data and the cloud service provider? The international nature of cloud infrastructures means that these issues require solutions which work across jurisdictions.

5.10 Near real time Cloud (e.g. Media transformation (rendering/transcoding))

One of the main benefits of cloud technologies is that of transparency; the end user neither knows nor cares where in the cloud his application executes as long as the SLAs are maintained. There are a number of existing and proposed consumer application areas such as media transformation where the use of cloud services can have a significant impact but the performance requirements are particularly demanding. New approaches to management may be required to address these requirements.

The Internet is increasingly used as a media delivery system for video and audio streams (both live and pre-recorded). There is an increasing interest in the consumer market in interactive media such as online gaming with dynamically rendered environments.

At the same time, the variety of terminal devices to consume this content is growing rapidly – from large screen (with 3D now emerging) displays connected to powerful computers and broadband networks to mobile phones with much smaller screens and much more constrained computing power and network connectivity. Transcoding of media to make it suitable for delivery to and display by such varied devices is challenging, particularly if it needs to be done in close to real time.

Such applications involve several cloud resources in their end-to-end operation, with implications for computing, storage and network performance. Consumers of networked media are geographically distributed and require encoding and delivery that is appropriate to their terminal equipment. Transcoding from an original base format to meet the needs of diverse users is computationally intensive. So too is dynamic rendering of shared virtual environments. Caching techniques or logistical networking may allow the results to be reused by many consumers, at the cost of additional storage. Multicast networking may also have a role. High definition or 3D TV or video generally involves very large volumes of data, presenting challenges for storage (both primary and staging) and network capacity. The inertia of data imposes severe restrictions on the feasibility of transcoding as a cloud service and further highlights the need to consider computing, storage and network resources together. Optimised media creation, adaptation, archiving, distribution and delivery based on the use of cloud infrastructure could be an attractive Platform-as-a-Service (PaaS) offering with scope for standardising interfaces to common features.

5.11 Software Licensing

Software licensing is a major inhibitor of the adoption of flexible computing models, including cloud infrastructure services. Cost savings in hardware, IT infrastructure management and energy can be negated by the need to purchase in advance, sufficient licences to cover the maximum size of an application deployment. The basis for software licensing varies considerably between vendors. Costs may be associated with the (maximum) number of servers, CPUs or cores the software is deployed on. Licence management systems are available to track and enforce this kind of usage. They may be related to the total or maximum concurrent number, of users. Less common arrangements include restrictions to a geographical area (e.g. on a single site) or to a limited set of named users. Also limiting flexibility are restrictions to running on specific physical hardware - requiring reactivation when transferred. Most of these licensing schemes are based on assumptions about the kind of physical infrastructure used to run the application in question. It is therefore not surprising that issues arise when new approaches to computing infrastructure, such as cloud, begin to emerge.

Software vendors are unlikely to change licensing models unless they see new market opportunities or face a competitive threat. Standardisation does not have an obvious role in their business decisions. However, technical solutions to licence management could increase the flexibility and the range of viable options for a software vendor. For example, the ability for software licensed to a particular user to run in an external cloud environment on behalf of that user would be desirable. This might include the ability to track concurrent use in both private and public cloud environments. Details of the use of licensed software and the mechanisms to audit and enforce licence terms in a cloud environment need to be specified in cloud service SLAs.

History

Document history		
V1.1.1	April 2010	Publication