The Asia Cloud Computing Association

Cloud Assessment Tool White Paper

www.AsiaCloud.org

2012 August 18

Version 3.0.5

ABOUT THE ASIA CLOUD COMPUTING ASSOCIATION

The Asia Cloud Computing Association (ACCA) fosters collaboration and innovation in Asia to drive adoption of cloud computing regionally. ACCA's outreach efforts extend to policy and regulation, security, best practices, and market education. Members include: Alcatel-Lucent, AT&T, Cisco Systems, Citrix, CITIC CPC, CloudGarage, Dimension Data, EMC Corporation, Equinix, Genetic Finance, Global Yellow Pages, Hong Kong Cyberport, Hiring Solutions, Huawei, iPerintis, Microsoft, NetApp, Nokia Siemens Networks, PLDT/Smart, Rackspace, Telstra Global, Telenor, TrustSphere, Verizon and Workday.

For membership information, visit: www.asiacloud.org

ABOUT THE AUTHORS

The development of the framework was a joint effort undertaken by the members of the Cloud Assessment Tool working group – which took into consideration both end user and service provider perspectives.

In particular, we are grateful for the valuable contributions and expertise of the following ACCA members: Johannes Prade (Chair-NSN), Michael Murphy (NSN), Joey Limjap (PLDT), Rolan Tanagras (PLDT), Wilhelm Paukert (PLDT), Simon Delord, (Alcatel-Lucent), Neo Teckguan (Huawei), Ray McQuillan (Terremark/Verizon), Syd Wong (Genetic Finance), Per Dahlberg (ACCA) and Mark Ross (ACCA).

Table of Contents

Introduction

The Cloud Assessment Tool (CAT) was developed by the CAT Working Group (WG) in the Asia Cloud Computing Association (ACCA). It was refined through extensive and in-depth discussions over a period of 2 years between members of the WG and by looking at relevant cloud and IT specifications.

The original mandate was to define the requirements placed on IaaS/PaaS solution providers to support stringent cloud applications. However, that perspective was subsequently extended to cover all application requirements. As such, its final realization has broad applicability.

It is unique and powerful, with immediate gains for those who adopt it. By publishing it here, we hope its usage will be broad and feedback will come, thus allowing us to improve it even further.

Note: The ACCA is in the process of developing an online version of CAT to further simplify its usage and understanding.

Executive Summary

Cloud computing is being used for an increasingly wide range of applications, some having demanding availability, security and performance requirements. Implementing such applications on clouds can be difficult because the operational criteria that characterize the behavior of clouds are complex and rarely well defined.

The framework developed by the ACCA for the Cloud Assessment Tool (CAT) addresses this issue. It does so by specifying operational criteria that should be well specified for such applications. Users benefit from the framework by having a pre-defined set of requirements that define their needs. Cloud providers benefit from the framework by having a standardized language to define their service offerings. When both come together, the result is that subjectivity and interpretation are eliminated in the mapping process. When the mapping is imperfect, the implication is that the onus is on the application to bridge any gaps, and this too is valuable information.

While the framework provides the highest value to stringent applications, it can also be applied to other application types. *Levels* were introduced to accommodate this, with each level representing a set of criteria appropriate for a particular class of applications.

The current variant of the framework has 172 criteria, divided into 8 categories, with 4 levels.

Work in progress

This iteration of the framework should be seen as a first step. We invite participation by industry stakeholders and interested parties to put forward suggestions that will help further

develop the framework.

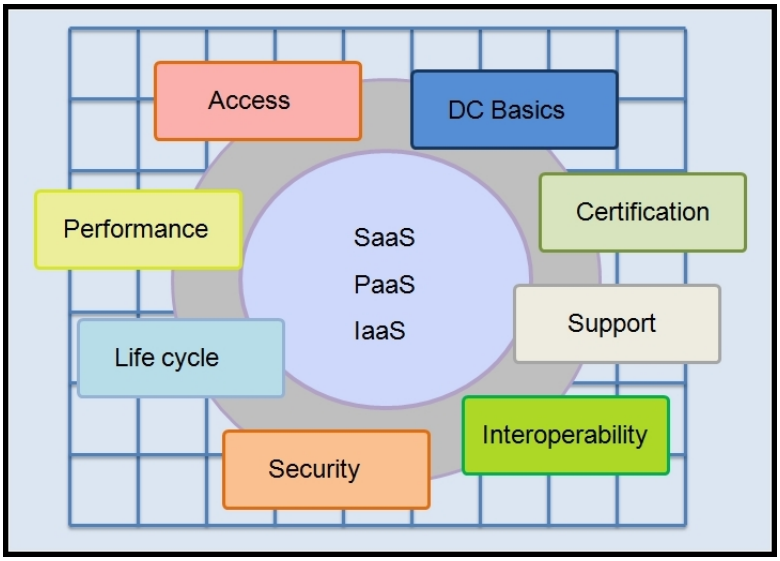
The Framework

The CAT framework is similar to that used by other industry forums and standards groups. It analyzes applications based on their needs, expressed as *criteria.* Criteria are divided into *categories*, for simplification. To support different types of application needs, *levels* are defined. Each level consists of a selection of criteria representing a similar class of applications. There are 8 categories, 172 criteria, and 4 levels.

Categories

The CAT framework starts with eight *categories* of broad operational performance characteristics of a cloud solution, as shown below. We would expect these to not change frequently.

Operational Categories



A brief description of each is as follows:

Security	Privacy, information security, regulatory
Life Cycle	Long-term support impacting customer business processes
Performance	Runtime behavior of deployed application software
Access	Connectivity between the end user and cloud service provider
DC Basics	Data Center physical infrastructure
Certification	Degree of quality assurance to the customer
Support	Deployment and maintenance of applications
Interoperability	Cloud hypervisor interfaces to applications

Levels

To support a range of applications, levels are defined across each category.

They are similar to the classification system used by the Uptime Institute which defines 4 data center models, referred to as Tiers I-IV. Tier I defines a data center with quite basic reliability, whereas Tier IV defines a data center having a highly redundant architecture.

The CAT framework uses a similar classification system to define 4 levels for each of the 8 categories of criteria. An example is shown here of a hypothetical cloud provider solution.

Levels Example

Categories	Level 1	Level 2	Level 3	Level 4
DC Basics			✔	
Access				✔
Performance			✔	
Life cycle			✔	
Security		✔		
Interoperability			✔	
Support	✔			
Certification & Compliance			✔	

The levels defined were based on an understanding of different application types, and loosely follow the concept of:

Level 1 Typical enterprise cloud solution
Level 2 Stringent application
Level 3 Telecommunications grade
Level 4 Beyond telecommunications grade

It is important to note that Level 4 is not *better* than Levels 3, 2 or 1. It is simply a matter of relevance whether a particular level is needed by an application or not. Also, an increase in level would generally imply a higher price from a cloud provider, and thus may not be appropriate.

As will be shown next, *levels*, by association, also apply to criteria.

Criteria

Across the 8 operational categories, 172 criteria have been defined, but this number could change over time as new requirements come to light.

Criteria can be measured, e.g. latency, or have a binary value, e.g. compliance or non-compliance with a global standard such as ISO 9000.

Criteria that cannot be determined precisely or are not applicable to all markets are not included in the framework. As they are by definition, imprecise, end users can interpret them according to the individual requirement of their applications and the different markets in which they operate.

Since criteria essentially express the details of a category, levels are made precise by the values of criteria. The following table illustrates this point by showing the 11 criteria for the Certification and Compliance category, mapped into 4 Levels.

Certification and Compliance Criteria Example

	PCI	SAS70	ISO 9000	ISO 27001/2	ISO/IEC 20000	EN16001 ISO50xx	Uptime Institute	HIPPA	Vendor cert	legal	patch
L4											
L3							Yes			Yes	
L2	Yes	SSAE16	Yes		Yes	Yes		Yes			
L1		Yes		Yes					Yes		Yes

The four 'Yes' in the Level 1 row indicate the minimum set of requirements needed to comply with this level. Five additional requirements are needed for Level 2 compliance, as well as the need to adhere to the stricter rules of SSAE16. Two additional criteria, "uptime" and "legal" are needed to reach Level 3.

Level 4 is not populated. This will allow for possible more strict requirements in the future.

All requirements stated for one level are carried forward to the levels above, where new requirements might be added, but the 'Yes' are not repeated in the diagram.

In many cases, criteria names are self-defining, but for clarity some brief definitions are provided below for the Certification and Compliance set:

PCI	Payment Card Industry
SAS 70	Standard for auditors to assess internal controls.
ISO 9000	Quality assurance
ISO 27xxx	Information Security Management
ISO/IEC20K	Follows IT service management 20000-1 and 2.
EN/ISO	Energy management

Uptime	Time a machine is operating or can be operated
HIPAA	Health Insurance Portability & Accountability Act of 1996
SW cert	Integrity of commercial software products
Legal obligation	Must follow legal obligations imposed.
SW security	Update offer when security leaks occur.

Detailed descriptions of all criteria can be found in the Appendix.

Matching User Needs to Cloud Provider Solutions

The framework described allows users to both define their needs and assesses cloud providers' offers. Indeed the highest value comes when the two come together. Let's look at a practical example.

In the first step, a hypothetical user employs the framework as a template to identify the requirements of an application. Consider the previous example shown illustrating the eight criteria and four levels.

In the second step, the user evaluates a cloud provider's solution using the same framework. An example is shown here for cloud provider (or vendor) "A."

When the levels of the cloud provider's solution are equal or higher than the requirement of the application then the offer is a good match.

	Vendor "A"
DC Basics	L1
Access	L1
Performance	L1
Life cycle	L3
Security	L1
Interoperability	L4
Support	L1
Cert. & comp.	L1
Average score	L1

If several offers are a good match, then the user will normally make a decision based on financial considerations. In scenarios where there is no match, or where the matching offer is too expensive, then the user would have to go back to the first step and see if there is an acceptable tradeoff or look at resolving disparities at the application level (vs. having the cloud provider resolve them).

Conclusions

As cloud computing evolves and is adopted by applications having different performance requirements than is the norm today, there is a need to define what requirements are important for the user and what capabilities are offered by the cloud provider.

The framework described above facilitates both of these activities, in a structured, objective way, thus simplifying the matching process between user and provider. It is, in essence, a tool, referred to as the ACCA Cloud Assessment Tool (ACCA CAT). An online version of it is being developed.

References

[1] http://uptimeinstitute.com/

[2] DMTF for OVF

[3] DMTF Virtualization Management (VMAN) (http://dmtf.org/standards/vman)

Appendix A - Categories

Categories are a means of specifying operational needs. There are 8 define in the CAT framework.

1. Certification & Compliance
2. Security
3. Lifecycle
4. Performance
5. Access
6. Data Center
7. Support
8. Interoperability

All application needs can be expressed as a collection of measureable details within one of these categories. Those measureable details are criteria. The Performance category, for example, is described by 8 criteria.

Levels are also applied to categories. For example, Level 2 of the performance category consists of all its criteria meeting a Level 2 standard. Two of the 8 performance criteria are *availability* and *VM to VM latency*. To meet a Level 2 standard for performance, availability must equal or exceed 99.99 and VM to VM latency must be equal or less than 10ms.

Appendix B - Levels

There are many levels of service provided by applications and cloud providers. To simplify and normalize those levels, 4 were defined in the CAT framework.

The decision to use 4 was to loosely align with the Uptime Institute's Data Center *Tiers.*

Levels apply to an entire application (or cloud solution), and to individual categories and criteria. It is essentially, a telescoping concept.

For example, for a cloud solution to meet a Level 2 standard, all categories must also meet a minimum Level 2 standard. For a category to meet a Level 2 standard, all criteria within that category must meet a Level 2 standard.

By defining the CAT Framework as such, applications can chose to define their needs in 3 different ways, in order of detail:

1. A Level 2 cloud solution
2. A Level 2 performance category, but a Level 3 security category, etc..
3. A Level 2 latency criteria but Level 3 availability criteria, etc..

This hierarchy of specifications supports the novice and expert approach, and thus makes the CAT framework very usable, but also very flexible, if needed.

While difficult, Levels were loosely intended to define a standard or grade of services as follows:

Level 1	Typical enterprise cloud solution
Level 2	Stringent application
Level 3	Telecommunications grade
Level 4	Beyond telecommunications grade

Appendix C - Criteria Descriptions

Criteria are the foundations of the CAT framework. In totality they express all the possible needs of an application and the possible measureable details of an offered cloud solution. The CAT framework has defined 172 criteria in total, divided across 8 categories.

Each criterion is atomic, in that, it cannot be broken up further.

Each criterion has a numeric value or a binary indication (e.g., meeting or not meeting a standard). There are a maximum of 4 values per criteria, that relate to Levels. For example, the *availability* criteria within the *performance* category, has values of three 9's, four 9's, five 9's or six 9's for Levels 1 through 4, respectively.

Category 1: Certification and Compliance

Eleven criteria have been identified as being representative of a wide range of applications from various industries. Some represent legal requirements for compliance with global standards, others are process or IT oriented. By clarifying which ones are supported by a cloud provider, users can be relieved of doing their own assessments.

Certification and Compliance

	PCI	SAS70	ISO 9000	ISO 27001/2	ISO/IEC 20000	EN16001 ISO50xx	uptime	HIPPA	Vendor cert	legal	patch
L4											
L3							Yes			Yes	
L2	Yes	SSAE16	Yes		Yes	Yes		Yes			
L1		Yes		Yes					Yes		Yes

PCI: Payment card industry (PCI) compliance is adherence to a set of specific security Standards that were developed to protect card information during and after a financial transaction. PCI compliance is required by all card brands.

SAS 70: Defines the standards an auditor must employ in order to assess the contracted internal controls of service organization, e.g. hosted data centers. SSAE16 is a newly introduced optional add on for higher standards.

ISO 9000: A series of standards, developed and published by the International Organization for Standardization that define, establish, and maintain an effective quality assurance system for manufacturing and service industries.

ISO 27001/27002: The objective of this pair of standards is to "provide a model for establishing, implementing, operating, monitoring, reviewing, maintaining, and improving an Information Security Management System".

EN 16001: Gives recommendations and guidance on energy management systems. It is designed to help improve energy efficiency by outlining how to implement processes that reduce greenhouse gas emissions. It enables the development of an energy policy and set objectives for the management system in line with legal and regulatory requirements.

ISO 50xxx: establishes a framework to manage energy for industrial plants; commercial, institutional, or governmental facilities; or entire organizations. Targeting broad applicability across national economic sectors, it is estimated that the standard could influence up to 60% of the world's energy use.

ISO 50001:2011 specifies requirements for establishing, implementing, maintaining and improving an energy management system, whose purpose is to enable an organization to follow a systematic approach in achieving continual improvement of energy performance, including energy efficiency, energy use and consumption

Uptime: The time during which a machine or piece of equipment is operating or can be operated.

HIPAA: Health Insurance Portability and Accountability Act of 1996.

SW vendor certification: Validates the integrity of commercial software products. It indicates the competence and ability of the provider to operate or offer any third party SW.

Legal obligation of provider: Providers are obliged to follow legal obligations imposed on them according to the jurisdictions they operate in. Since this may impact the user, who may not even be aware of it, the proposal is to "advertise" these proactively.

SW security patch process: The provider is requested to update offer with respect to SW security leaks. This may have impact on the user in requires an appropriate, well defined interaction.

ISO/IEC 20000: The provider indicates that they have followed the standardized methods of IT service management, particularly 20000-1 and 20000-2 in revisions from 2011 and 2012 respectively.

The definitions and taxonomy for the remaining seven criteria can be found in the appendix of this document or downloaded from the ACCA website.

Category 2: Security

Security currently is one of the hottest topics in the cloud-computing ecosystem. Virtually all stakeholders, including governments, have identified the need to provide guidance and regulation. This has also been recognized by ACCA, which has responded by forming a working group dedicated to cloud security issues. In the context of the CAT, we limit this category to aspects that would have an immediate impact on the interaction between user and provider or application and resources.

We recommend also to consult the Cloud Security Association's (CSA) guidelines for an in depth understanding of this topic. National regulations have to be considered as well, which are not addressed here.

Security

	Authen-tication	User account logging	Role based access ctrl	Pro-tection	Loss	Data removal	Data encrypt-tion	Location aware-ness	User def sec dom	IP-spoofing	Fire-walls	Sec inc rep & mgmnt
L4					Business Continuity Process							
L3	Two-factor authentication				DR + comprehensive backup		Private key encryption			Intrusion detection	User configurable	Yes
L2		User a/c mgmt			Backup snapshot					Yes		
L1	Standard authentication	Self serve access, logging & tracking	Manage diff access rights	No access to s/w, data by other users	Retrieve and restore data and s/w	Entire and irretrievable removal of data at user request		Known and/or specified by user	Yes		Yes	

Authentication: Level 1: Standard methods to authenticate the portal as well as the API access by a user. L3: For mission and business critical use cases the advanced two-factor authentication is required.

User account logging: Self serve access and maintenance requires logging of user access and user actions for tracking. L2: Added capability of logging management activities on requested resources and other actions.

Role based access control: In order to mirror business processes even in cloud environments, managing the access for several members of an organization on the same data (user account) with different access rights is required.

Protection: Assurance that software, computing results, data and etc., cannot be accessed or infringed upon by other users. This includes inter-virtual machine attack prevention, storage block level isolation and hypervisor compromise protection.

Loss: In this context loss means that the user can retrieve the deposited assets like

software, data and etc., even though the service is out of operation. The redundantly stored information must be monitored for integrity. It is also expected that the software and data will be restored automatically after an entire system failure. L2: User initiated backups and snapshots enable user to restore to most recent version. L3: Disaster recovery protection and comprehensive backup and restore services added. L4: Provider implements well-defined methods for users to establish a business continuity process.

Data removal: In case a user requests his software or data to be deleted, all data/software stored in the cloud must be entirely and irretrievably removed. This requires that the appropriate techniques be employed to locate the data and all its backups, encrypted or otherwise, and to completely erase all of them into an unrecoverable state.

Data encryption: Even if data is protected against access by other cloud users, encryption of stored data might be necessary to avoid unauthorized read of data through cloud provider employees, tools and etc. L3: the cloud provider establishes a customer's individual private key encryption mechanism.

Location awareness: The user receives an indication of where his data is being stored and processed. User can specify where software and data have to be stored, run and processed. Provisions are in place to ensure all data and backups are stored only in these locations agreed by contract or the service level agreement.

User defined security domains: Different application types require different protection levels. Therefore, deploying SW into a cloud requires grouping of SW parts into domains, e.g. DMZ, application tier, data base tier, for which appropriate filter rules have to be defined and applied (visibility of a domain).

IP spoofing: L2: Any attempt to maliciously access an application from within the cloud should be blocked. L3: provider implements proactive measures for intrusion detection.

Firewalls: Resources assigned to a user should have protection against external attacks. L3: Firewall allows user specific adjustments for highly critical applications.

Security incident management and reporting: L3: Besides protecting against attacks like Denial of Service, it is desirable that a provider detects and handles security attacks from outside as well as from inside the cloud. An immediate indication/reporting to users is needed to quickly react on an individual basis.

Category 3: Lifecycle

Cloud computing is a way for users to outsource selected IT functions, infrastructure, business apps and etc. Even when "in the cloud", these functions remain part of an enterprise's business processes. In addition, these functions are not static; they need extensions updates, changes and etc. Since an enterprise may use different providers for cloud services, it is important to address these life cycle and process related aspects in a well-defined standardized way.

Lifecycle Criteria

	Dev. Roadmap	Security management	Config mgt	service mgt	Reporting	Portal	Deployment	Billing	Ticketing
L4	CMM L4	24x7 Security Op Ctr		ITIL L4/5					
L3	CMM L3		Compre-hensive	ITIIL L3	Compre-hensive rep		Auto VM upgrade	History	View entire ticket DB and stats
L2	CMM L2	Crisis management process		ITIL L2	Dashboard	Service catalog	Custom image support		Problem logging on request
L1	Yes	Proactive security monitoring	Basic	SLA commitment	Basic rep	Self service	Aut-prov		

Development Roadmap: Ensure that the service provider has a planned way forward (process) to evolve available features and introduce new capabilities. L1: Process available. For L2 to L4 a well-defined approach comparable to “Capability Maturity Model” is needed, e.g. L2 is CMM L2, L3 is CMM L3 and L4 is CMM L4.

Security management: In addition to the “technical” aspects of security, which are covered in the “Security” category, this topic here covers the procedural aspect of it. L1: Provider performs proactive security monitoring. L2: calls for a crisis management process to be established. L4: Comprehensive supervision by 24x7 security operations center.

Configuration management: L1: Basic configuration management refers to the support of the configuration management database (CMDB) capturing basic configuration items (CI) that allow the on-line tracking of the cloud resources subscribed by the end-users. L3: This refers to the support of an enterprise class CMDB capturing comprehensive CI information with correlation capability between CI that allow the on-line track of the cloud resources and CIs relationships subscribed by the end-users.

Service Management: In general terms cloud services are IT services remotely offered to the customer. There are well defined and well structured methods available to determine how services are expected to be managed in an enterprise environment. One should expect the same structured approach to IT service offerings from a service provider as would be expected from an in-house IT organization. Cost, effort and rigidness increase with the ITIL level and present a natural way for mapping it into the CAT framework levels.

Reporting: L1: Basic reporting refers to the support of simple on-line reports that cover essential cloud services. Basic reports may include report on Cloud resources subscribed for VMs, Storage, Internet IP, Bandwidth usage, etc. L2: In addition to L1 requirements an on-line information dashboard should be made available to the user, showing the list of essential Cloud services currently being deployed and utilized. It may include real-time update of information on status of VMs, Storage usage, Storage Buckets, Data transfer and others. L3: Comprehensive reporting extends L2 capabilities to include pre-defined reports and customer customizable reports that users can receive via various channels including on-line viewing, download (as .pdf, .html, .doc, etc.) from web portal or via e-mail. Comprehensive reports may include report on Cloud resources subscribed for VMs, Storage, Internet IP, bandwidth usage, VM’s performance, network performance, security threats, billing and invoicing, ticketing statistics, etc.

Portal: L1: "Self Service" is the cloud Web portal feature that enables customers to perform most of the essential services themselves. This includes provisioning resource, managing resources such as controlling VM status (reboot, shutdown, restart, etc.), viewing various subscribed services, downloading essential support documents (e.g. user guides and FAQ list, etc.). L2: "Service Catalog" is a function in the cloud Web portal that supports listing and grouping of all available products and services to which customers can browse and subscribe.

Deployment: L1: Auto-provisioning is the ability of the cloud Web portal to capture customer requested cloud service subscription and automatically provision the cloud services without human intervention. L2: Custom image support permits users to customize a VM server image and quickly redeploy that via a portal on new shared cloud servers. L3: Auto VM-upgrade permits a user to upgrade the virtual core, RAM and Storage of a VM from small configuration to higher configuration without migrating the operating system and data.

Billing: L3: Service provider keeps a history of the customer's use of chargeable resources and services.

Ticketing: L2: Problem logging supports submission of problems and requests from a customer via a Web form, e-mail link or other methods. L3: ticketing functions allow all customer submitted requests and problems to be logged and retrieved for viewing by customers along with statistics.

Category 4: Performance

Designing and running applications with varying levels of availability and performance in an IaaS environment requires visibility of and control over the deployed resources. The performance parameters and values below reflect what is typical of telecom and data base applications.

Performance Criteria

	Availability %	VM to VM latency	Scalability	Elasticity	Redun-dancy	Load Distribution	Control	Storage
L4	99.9995	100us			Cluster deployment API		SLA for DC LAN	
L3	99.999	1ms	+autoscaling support for application	+ >1 sec to start/end new VM	HA-database as part of offering	Selectable distribution criteria	Pinning of vCPU to logical core	Tiered storage
L2	99.99	10ms	+ resource utilization monitoring tool	+ programmatic interface	+ auto restart of dropped app	Load distr. redundancy	Multi VM to VM L2 networks	
L1	99.95	100ms	Flexible granularity of resource scaling	Scale up and down	Availability zone	Intell load distr + progr i/f to load distri events		

Availability: Refers to the length of time the service is offering without interruption (outside defined maintenance windows). L1: 99.95% represents standard IT hardware and software

runtime uptime. L3: 99.999% is the widely accepted requirement in the telecommunications industry for high-availability services. L4: 99.9995% is required to support highly critical business applications (e.g. database applications).

Inter-VM latency: Refers to the maximum time allowed to send a message from one software instance to another running on different physical machines. L1: 100ms (e.g. for basic Web apps). L2: 10 microseconds (for distributed telecommunications DB services). L3: 1 microseconds (financial trading). L4: 100 nanoseconds (for at least 99.5% of messaged, for real time apps highly sophisticated redundancy architectures).

Scalability: Measures the performance level to which a deployed application can be engineered. L3: Auto scaling support to scale an application without user monitoring or intervention.

Elasticity: Addresses the how fast a deployed application can increase its performance response with increasing service requests. L3: elapsed detect-trigger time between application/host and infrastructure management.

Redundancy: Redundancy architectures frequently rely on a well-defined set of software components to preserve states and transactions. L3: Service incorporates High Availability-database for high availability software design. L4: Means for cluster deployment, ensuring that software units are deployed on the physical infra structure according to requirements specified by the application.

Load distribution: Widely used scheme to implement scalable and reliable services.

Deployment Control: Highly reliable applications implement redundancy schemes, which must be preserved when deploying to IaaS. L1: Resources are fully transparent to user. L2: NaaS (Network as a Service) for multiple networks between virtual machines. L3: Application performance engineering requires dedicated CPU resource assignment -> pinning of vCPU to cores. L4: SLA for e.g. delay, jitter, throughput, availability

Storage: L3: Deploying critical and highly available applications to an IaaS offering requires special visibility of the storage capabilities so that the application architecture can be adjusted accordingly.

Category 5: Access

Generally speaking, the responsibility of a service provider is limited to the service provider's own network over which he/she yields control. However, with cloud computing, where one of the basic principles is the remote access to the service by the end-user (regardless of who owns the links and network(s) between the user and the service), this concept is challenged. Typically, "public" Internet access is the standard way to connect to the cloud service. For critical services, a dedicated "private" link (physical or virtual) is required to ensure security and throughput. We highlight three aspects for consideration.

1) For public access, the cloud service provider should clearly describe available

service levels on his/her own network.

2) For public access, the user has to check and reach separate agreements with other network access providers in the chain.

3) For private access, there will be an organization responsible for this dedicated connectivity, which also controls the path between the user and the cloud provider. In this case, SLA requirements can be specified and possibly agreed upon.

Note: By definition, there is an unknown distance and space between the user and the provider. Therefore, a topological view will be required.

Access Criteria

					QoS	QoS	QoS
	Access	Availability	Scalability	Reliability	Class of Service	Packet Loss	Delay
L4		100%				0.001%	
L3	Public / Private On-demand	99.9999%	Selectable bandwidth			0.01%	<=180ms
L2	Public / Private	99.996%					
L1	Public	99.99%	Min bandwidth guarantee	<=50ms	Non-mission critical, mission critical, l-time voice/video	<=0.1%	<=230ms

Access: This parameter measures the type of access. L1: Access through public Internet. L2: a) Enhanced availability public Internet access. B) Private, i.e. dedicated link access, e.g. through VPN and other methods. L3: enhanced public access or on-demand private access.

Availability: Indicates the "guaranteed" level of uptime of the network access.

Scalability: Capability to increase and decrease user's access bandwidth based on actual capacity demand.

Reliability: Reliability reflects the ability to rapidly detect and recover from link failures. This parameter cannot be assessed in the case of public internet access (i.e. best effort) since there is no organization involved in the business relation between cloud user and provider. In case of "private access", this parameter would be relevant for certain types of connection.

Quality of Service: Similar to the "Reliability" parameter above, the Quality of Service (QoS) parameter is difficult to assess for public access, as it is on a "best effort" basis. The QoS parameters reflect the features, which influence the quality of delivery of a service. The values of the parameters are according to standard's definitions (here 3GPP) with respect to certain use cases. The levels reflect the increasing requirements with respect to real time experience.

Delay: is the commitment to end-to-end latency. Packet loss: a guarantee to deliver

customer packets, expressed as the ratio of undelivered packets to total number of customer packets received by the network. Service class: allows priorities to be differentiated and data flows to be shaped for comparable services (non-mission critical, mission critical, real time).

Category 6: Data Center

The level of reliability of a cloud service provider may also be gauged based on the data center or geographic zone on which its infrastructure is located. The various tiers of data center types from which a cloud service is hosted have been used to map the parameters for the data center category. The horizontal axis represents the primary data center that hosts the cloud service. The vertical axis represent the secondary data center that would offer redundancy to the cloud service being offered.

L4: This is the highest level for clouds and can be considered as the most robust and least prone for failures. This cloud is hosted primarily in a Tier 4 data center and backed up by either a Tier 4 or Tier 3 data center.

L3: This cloud is hosted in two Tier 3 data centers OR a combination of Tier 4 and Tier 2 data centers.

L2: This cloud is hosted in either a Tier 4 / Tier 1 data center OR Tier 3 / Tier 2 data center combinations. Local redundancy may also be considered when hosted in a Tier 4 data center.

L1: This is the lowest level and has higher risks of failure. This cloud is hosted in two Tier 2 data centers OR a combination of Tier 3 and Tier 1 data centers. Local redundancy may also be considered when hosted in a Tier 3 data center.

Data Center Criteria

Primary / Secondary	Tier 4	Tier 3	Tier 2	Tier 1
Tier 4	L4	L4	L3	L2
Tier 3	L4	L3	L2	L1
Tier 2	L3	L2	L1	Non-carrier Grade
Tier 1	L2	L1	Non-carrier Grade	Non-carrier Grade
Local redundancy	L2	L1	Non-carrier Grade	Non-carrier Grade

Data Center Tiers are nothing but standardized methodology published by the Uptime Institute to evaluate the quality and reliability of a data center's server hosting ability. This

rating system has four levels starting with Tier 1 as the data center with the most basic requirements and ends with Tier 4 offering the highest availability.

Tier 1 Data Centers have non-redundant capacity components and a single, non-redundant uplink to the server guaranteeing 99.671% availability.

Tier 2 Data Centers have redundant capacity components and a single, non-redundant uplink to the server guaranteeing 99.749% availability.

Tier 3 Data Centers are concurrently maintainable data centers with dual-powered equipment and multiple independent up-link to the server guaranteeing 99.982% availability.

Tier 4 Data Centers are fault tolerant data centers with dual-powered equipment, continuous cooling systems and multiple, independent and physically isolated systems that provide redundant capacity components and multiple, independent, active uplinks to the server guaranteeing 99.995% availability.

Category 7: Support

NIST characterizes cloud computing as a "ubiquitous" and "on demand self-service" that can be rapidly provisioned and released "with minimal management effort or service provider interaction." Enterprises, big and small, will still require support from the cloud service provider, especially for business critical services. This is an aspect that has to be considered in the context of business continuity.

Support Criteria

	Customer Support	Service Responsiveness	Service Escalation	Monitoring, Audit, Management	MTTR*	Change Management support	Incident Response Time (Pri1)	Service Training
L4	Assist trouble shooting			Assisted Montoring				
L3	Engineering Support				2 hours	Customer initiated change requests	15 minutes	
L2	Service Desk	<=20s		Incident management for all services				
L1	Portal	24/7	4 levels	100% availability for customer's access	4 hours	Provider objectives	30 minutes	Yes

Customer support: Methods and capabilities available for how a user can interact with the service provider. L1: Any relevant interaction through portal. L2: Human interaction possible through service desk. L3: application architecture and performance engineering support. L4: providing troubling shooting support for customer SW.

Service responsiveness: Time it takes for the service provider to respond to calls or customer inquiries. L1: 24/7, customer can call all around the clock. L2: Customer request

accepted with in 20s. L4: Customer will immediately reach the relevant service agent.

Service escalation: The levels of escalation provided to customer for proper trouble handling.

Monitoring/audit/management: Cloud implies “self-service” and “ubiquitous”. Therefore, the user needs to be able to monitor and manage the behavior of the deployed user objects at any point in time from any location. L1: Permanent access to monitor and manage the user’s objects. L2: Enhanced by incident management of all services of the provider. L4: Provider assisted monitoring.

MTTR: mean time to repair is the time it takes for service provider to correct reported faults or incidents concerning services acquired by user. L1: 4 hours. L3: 2 hours.

Change management support: An acquired service is neither static nor unchanged at all times, it may need internal repairs, upgrades, etc. This impacts the behavior of the user’s service or application. L1: Well defined change management for service provider’s objects. L3: User can request changes.

Incident response time: Maximum time it takes for service provider to react and act on Priority 1 incidents (an event where a service/application is not working or accessible). L1: 30 minutes; L3: 15 minutes.

Service training: Structured training and transfer of know-how is important for businesses.

Category 8: Interoperability

Interoperability reflects the possibility to aligning IT hardware and software systems with the help of open standards and interfaces facilitate communications and the exchange of data. In the context of cloud computing, interoperability describes the level of cooperation capability a cloud service can reach with other cloud services.

The higher the level of cooperation, the more cost efficient a business process can be established leveraging cloud services from various providers.

This category together with the “life cycle” and the “compliance” categories address the operational aspects when utilizing cloud services and, therefore, have impact on the user’s operating expenses.

Note: Portability (i.e. the requirement to be able to run software on more than one computer system) is not explicitly addressed in this framework. Nonetheless, in the context of cloud computing, interoperability parameters help to support portability.

Interoperability Criteria

	Download Format	API	Access Device	Life Cycle	Monitoring Data	IDM	Virtual Management
L4							
L3		Non provider specific	Any browser	DMTF	Complex Monitoring tools		
L2	OVF		Open client			Standard	DMTF
L1		Provider specific	Customized				

Download format: This represents the format of the objects to be transferred into the provider's infrastructure. L2: DMTF [2] has developed a standard format called "open virtualization format" (OVF) which eases the federation across several offerings.

API: L1: The programmatic interfaces to request and monitor resources, to deploy and manage applications, mirror the feature differentiation as well as the fast innovation cycles of the service offerings. L3: Standardized API will reduce complexity and cost on the user side whenever several cloud services have to be combined or utilized.

Access device: L1: Vertically integrated service solutions may benefit from specialized or customized access devices. L2: For consuming any application as a service, standard methods for access are needed ("open client"). L3: For anywhere (e.g. mobile) and any device (e.g. BYOD) access a fully open access (any browser) is needed.

Life cycle: The user needs to integrate the requested services into the IT management processes and applications/services. L3: DMTF developed recommendation [3] allowing consistent life cycle management of different services.

Monitoring data: Any applications deployed in IaaS or PaaS environments must be monitored and managed by the user. L3: For business critical applications a comprehensive set of data (and appropriate monitoring capabilities) should be provided in standardized formats to enable import in user's monitoring tools.

IDM (Identity Management): Service providers must have a way of handling the identity of the user and the user must integrate the IDM of the outsourced service into their IT infrastructure. L2: There are many ways to address IDM. For interoperability reason an IDM system based on available standards is recommended.

Virtualization environment management: The user may need to integrate the requested resources into their IT management processes (especially for IaaS). L2: DMTF developed recommendation allowing consistent integration of several services.