International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.3510

(02/2016)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

Cloud Computing

Cloud computing infrastructure requirements

Recommendation ITU-T Y.3510

International Telecommunication Union

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Enhancements to NGN	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999
FUTURE NETWORKS	Y.3000–Y.3499
CLOUD COMPUTING	**Y.3500–Y.3999**
INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES	
General	Y.4000–Y.4049
Definitions and terminologies	Y.4050–Y.4099
Requirements and use cases	Y.4100–Y.4249
Infrastructure, connectivity and networks	Y.4250–Y.4399
Frameworks, architectures and protocols	Y.4400–Y.4549
Services, applications, computation and data processing	Y.4550–Y.4699
Management, control and performance	Y.4700–Y.4799
Identification and security	Y.4800–Y.4899

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.3510

Cloud computing infrastructure requirements

Summary

Recommendation ITU-T Y.3510 provides requirements for cloud computing infrastructure; these include the essential capabilities for processing, storage and networking resources, as well as the capabilities of resource abstraction and control.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y.3510	2013-05-22	13	11.1002/1000/11918
2.0	ITU-T Y.3510	2016-02-13	13	11.1002/1000/12713

Keywords

Capability, cloud computing, control, infrastructure, networking, processing, resource abstraction, storage.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T Y.3510

Cloud computing infrastructure requirements

1 Scope

This Recommendation identifies requirements for cloud infrastructure to support cloud services.

The scope of this Recommendation includes:

– an overview of cloud computing infrastructure;
– requirements for processing resources;
– requirements for networking resources;
– requirements for storage resources;
– requirements for resource abstraction and control.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T X.1601] Recommendation ITU-T X.1601 (2015), *Security framework for cloud computing.*

[ITU-T Y.3500] Recommendation ITU-T Y.3500 (2014) | ISO/IEC 17788:2014, *Information technology – Cloud computing – Overview and vocabulary*.

[ITU-T Y.3501] Recommendation ITU-T Y.3501 (2013), *Cloud computing framework and high-level requirements.*

[ITU-T Y.3502] Recommendation ITU-T Y.3502 (2014) | ISO/IEC 17789:2014, *Information technology – Cloud computing – Reference architecture.*

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 cloud computing [ITU-T Y.3500]: Paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand.

NOTE – Examples of resources include servers, operating systems, networks, software, applications, and storage equipment.

3.1.2 cloud service [ITU-T Y.3500]: One or more capabilities offered via cloud computing invoked using a defined interface

3.1.3 cloud service customer [ITU-T Y.3500]: Party which is in a business relationship for the purpose of using cloud services.

NOTE – A business relationship does not necessarily imply financial agreements.

3.1.4 cloud service provider [ITU-T Y.3500]: Party which makes cloud services available.

3.1.5 emergency telecommunications (ET) [b-ITU-T Y.2205]: ET means any emergency related service that requires special handling from the NGN relative to other services. This includes government authorized emergency services and public safety services.

3.1.6 emergency telecommunications service (ETS) [b-ITU-T E.107]: A national service providing priority telecommunications to ETS authorized users in times of disaster and emergencies.

3.1.7 logical resource [b-ITU-T Y.3011]: An independently manageable partition of a physical resource, which inherits the same characteristics as the physical resource and whose capability is bound to the capability of the physical resource.

NOTE – "independently" means mutual exclusiveness among multiple partitions at the same level.

3.1.8 management system [b-ITU-T M.60]: A system with the capability and authority to exercise control over and/or collect management information from another system.

3.1.9 virtual resource [b-ITU-T Y.3011]: An abstraction of physical or logical resource, which may have different characteristics from the physical or logical resource and whose capability may not be bound to the capability of the physical or logical resource.

NOTE – "different characteristics" means simplification or extension of the resource characteristics. "different characteristics" allows the virtual resource to expose access or control methods different from the original physical or logical resource.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 hypervisor: A type of system software that allows multiple operating systems to share a single hardware host.

NOTE – Each operating system appears to have the host's processor, memory and other resources, all to itself.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CPU Central Processing Unit

CSC Cloud Service Customer

CSP Cloud Service Provider

DFS Distributed File System

DHT Distributed Hash Table

DNS Domain Name System

ET Emergency Telecommunications

ETS Emergency Telecommunications Service

I/O Input/Output

iSCSI Internet Small Computer System Interface

LAN Local Area Network

NAS Network Attached Storage

NFS Network File System

NTP Network Time Protocol

OS Operating System

QoS Quality of Service

SAN Storage Area Network

SLA Service Level Agreement

vCPU virtual CPU

VI Virtual Infrastructure

VM Virtual Machine

VPN Virtual Private Network

5 Conventions

In this Recommendation:

The keywords "is required" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "is prohibited" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus this requirement need not be present to claim conformance.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

6 Overview of cloud infrastructure

In this Recommendation, cloud infrastructure includes processing, storage, networking and other hardware resources as well as software assets.

Abstraction and control of physical resources are essential means to achieve the on-demand and elastic characteristics of cloud infrastructure. In this way, physical resources can be abstracted into virtual machines (VMs), virtual storages and virtual networks. The abstracted resources are controlled to meet cloud service customers' (CSCs) needs.

The main characteristics of cloud infrastructure are:

– Network centric: The cloud infrastructure consists of distributed resources including processing, storage and other hardware resources that are connected through the networks;

– On-demand resource provisioning: The cloud infrastructure dynamically provides resources according to CSCs' needs;

– Elasticity: The cloud infrastructure is capable of expanding or reducing its resources to accommodate the current workloads;

– High availability: The cloud infrastructure is capable of providing required resources under the conditions stated in the service level agreement (SLA);

– Resources abstraction: The underlying resources of cloud infrastructure (processing, storage, networking, etc.) are invisible to the CSCs.

NOTE – For high level cloud computing requirements, please refer to [ITU-T Y.3501].

7 Requirements for processing resources

Processing resources are used to provide essential capabilities for cloud services and to support other system capabilities, such as resource abstraction and control, management, security and monitoring.

The basic unit of allocation and scheduling of processing resources is a computing machine. A computing machine can be physical or virtual. The capability of a computing machine is typically expressed in terms of configuration, availability, scalability, manageability and energy consumption.

7.1 Physical machine requirements

The physical machine requirements include:

- It is recommended to support hardware resource virtualization.
- It is recommended to support horizontal scalability (e.g., adding more computing machines) and vertical scalability (e.g., adding more resources with a computing machine).
- It is recommended to use power optimization solutions to reduce energy consumption.

7.2 Virtual machine requirements

The virtual machine provides a virtualized and isolated computing environment for each guest operating system (OS).

The virtual machine requirement includes:

- It is required to support migration of virtual machines between different physical computing machines.

7.2.1 CPU virtualization

The central processing unit (CPU) virtualization allows running multiple virtual CPUs (vCPU) on a single physical CPU.

The CPU virtualization requirement includes:

- The virtual machine's vCPUs' computing capability can optionally be specified as a fraction of a physical CPU.

7.2.2 Memory virtualization

The memory virtualization includes memory allocation at the start of a virtual machine, memory utilization monitoring during virtual machine operation and memory release at the shutdown of a virtual machine.

The memory virtualization requirement includes:

- It is recommended that while a virtual machine is active, the hypervisor monitors memory usage and reallocates unused memory to other virtual machines dynamically.

7.2.3 Input/Output device virtualization

The Input/Output (I/O) virtualization allows division of the physical I/O device into several logical instances for use by different virtual machines.

The I/O device virtualization requirements include:

- It is required for the hypervisor to support I/O virtualization capabilities.
- It is required that a virtual machine is capable of using virtual I/O devices abstracted from the physical I/O devices.
- The number of virtual I/O devices is prohibited from being constrained by the number of physical I/O devices.

- The data of one virtual machine transferred through a shared physical I/O device is prohibited from being exposed to other virtual machines.
- Physical I/O devices can optionally be shared by multiple virtual machines.

7.2.4 Network interface virtualization

Network interface virtualization allows creating and deleting a virtual network interface for a guest virtual machine OS regardless of the number of physical network interfaces.

The network interface virtualization requirements include:

- It is recommended that a physical network interface can be virtualized into multiple virtual network interfaces.
- It is recommended that the virtual network interfaces from different virtual machines can be grouped into one virtual local network.

7.2.5 Duplication of virtual machine

The duplication of a virtual machine allows creating new virtual machines and virtual machine backup in the execution environment.

The duplication of a virtual machine requirement includes:

- A virtual machine can optionally be duplicated to create a new virtual machine with the same configuration.

7.2.6 Dynamic migration of virtual machine

The dynamic migration of a virtual machine is designed to provide service continuity and reliability dynamically.

The dynamic migration of virtual machine requirements include:

- It is required that the network configuration of migrated virtual machines remain unchanged after migration.
- It is recommended that cloud service providers (CSPs) support the dynamic migration of a virtual machine.

7.2.7 Static migration of virtual machine

The static migration of a virtual machine means moving the virtual machine between different physical machines which results in the operating system rebooting.

The static migration of a virtual machine requirement includes:

- It is required that CSPs support static migration.

7.2.8 Management automation

The management system may perform operations such as starting or stopping a virtual machine, rebooting a server and applying software updates automatically.

The management automation requirement with regard to virtual machines includes:

- It is recommended that CSPs automate provision, activation, deactivation and other operations over the lifetime of virtual machines.

7.3 Software resources provisioning requirements

The software resources include the software for building cloud infrastructure resource pools and the software in support of service implementation.

7.3.1 Automated provisioning and deployment

The automated provisioning and deployment of software resources can reduce provisioning time and the workload for deployment.

The automated provisioning and deployment requirements include:

- It is recommended that software resources (e.g., executable files, drivers, libraries, documents, icons, etc.) are packaged into encapsulated files, which can be provisioned and deployed automatically.
- It is recommended that software resources be automatically provisioned and deployed to target devices or platforms without operator intervention.

7.3.2 Unified software resource management

Unified software resource management includes capabilities for licence information registration, allocation, recovery, expiration notification and metering.

The unified software resource management requirement includes:

- It is recommended that CSPs manage software licences in a unified manner.

7.4 Time-sensitive services requirements

Time-sensitive services (e.g., real-time communications using voice and video media) requirements include:

- It is required to prioritize resource allocation to time-sensitive processing.
- It is required to apply clock settings best practices (e.g., based on the network time protocol (NTP) [b-IETF RFC 5905]).

8 Requirements for networking resources

Typically, there are several types of networks involved in cloud computing services delivery and composition, such as the intra-datacentre network and inter-datacentre network, as well as the access and core transport network, etc.

To illustrate the cloud computing network concepts described in this Recommendation, a generic network model, which supports cloud computing infrastructure, is shown in Figure 8-1.

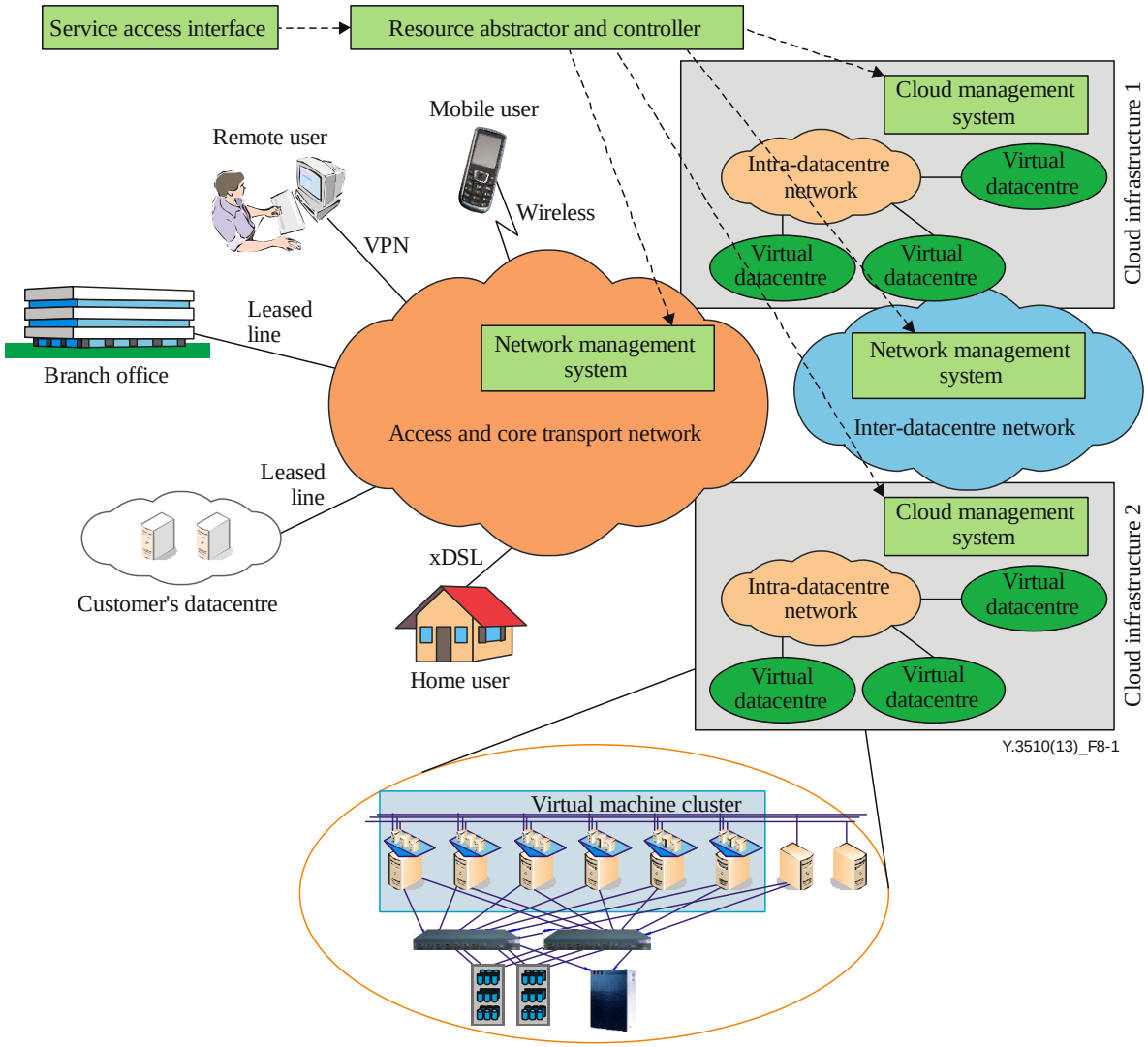

Figure 8-1 – Generic network model for cloud infrastructure

The generic network model shown in Figure 8-1 consists of the following blocks:

1) **Intra-datacentre network**: The network connecting local cloud infrastructures, such as the datacentre local area network used to connect servers, storage arrays and L4-L7 devices (e.g., firewalls, load balancers, application acceleration devices).

2) **Access and core transport network**: The network used by CSCs to access and consume cloud services deployed by the CSP.

3) **Inter-datacentre network**: The network interconnecting remote cloud infrastructures. These infrastructures may be owned by the same or different CSPs; an inter-datacentre network primarily supports the following two scenarios:

   – **Workload migration**, which means moving workloads from an enterprise datacentre to a CSP datacentre, or moving workloads from CSP to CSP (for resilience and maintenance).

   – **Server clustering** which allows transactions and storage replication for the business continuity.

Examples of inter-datacentre network models include:

1) private cloud datacentre to private cloud datacentre

2) private cloud datacentre to CSP datacentre

3) CSP datacentre to CSP datacentre.

NOTE 1 – For a description of a private cloud, please refer to [ITU-T Y.3500].

A centralized resource abstraction and control ensures the overall management of the cloud environment with:

a) Network management systems that are dedicated to network service providers. The processes supported by network management systems include management and maintenance of the network inventory and the configuration of network components, as well as fault management.

b) Cloud management systems are dedicated to CSPs. Cloud management systems support processes for maintenance, monitoring and configuration of cloud infrastructure resources.

NOTE 2 – Requirements for resource abstraction and control are provided in clause 10.

8.1 General requirements for networking resources

General requirements provided in this clause apply to networking resources of the access and core transport networks, intra-datacentre networks as well as inter-datacentre networks.

The general requirements for networking resources include:

- Networking resources (e.g., bandwidth, number of ports, network addresses) are required to be scalable;
- Networking resources are required to ensure services' performance and availability in order to meet SLA objectives;
- Networking resources are required to be able to adapt dynamically to the traffic generated by cloud services;
- Networking resources are required to support IPv4 and IPv6;
- Networking resources are recommended to support policy based control on flow by flow basis in a fine-grained manner.

8.2 Access and core transport network

The access and core transport network is used to connect the CSC to the CSP for the use of cloud services.

The access and core transport network requirement includes:

- It is recommended that the access and core transport network supports the delivery of cloud services in an optimal way in terms of performance, scalability and agility (e.g., through network programmability).

8.3 Intra-datacentre network

The intra-datacentre network is used to connect local datacentre cloud infrastructures, such as servers, storage arrays and L4-L7 devices (e.g., firewalls, load balancers, application acceleration devices).

The intra-datacentre network requirements include:

- The intra-datacentre network is recommended to provide appropriate means to cope with flexible network address space demands.
- The intra-datacentre network is recommended to provide elastic addressing for multi-tenant users.
- The intra-datacentre network is recommended to support different security policies for particular virtual machines.

- The intra-datacentre network is recommended to support different QoS policies for particular virtual machines.
- The intra-datacentre network is recommended to support the dynamic migration of virtual machines.
- The intra-datacentre network is recommended to support the traffic monitoring among virtual machines and network ports if needed.
- The intra-datacentre network is recommended to be able to provide multi-paths for particular multi-tenant users.
- The intra-datacentre network is recommended to support the establishment of a logical network among virtual machines.
- The intra-datacentre network is recommended to support public IP address and private IP address mapping.
- The intra-datacentre network is recommended to support dynamic DNS and static DNS for multi-tenant users.
- The intra-datacentre network is recommended to support network services (e.g., firewall, load balancer, virtual private network (VPN) services) for multi-tenant users.

8.4 Inter-datacentre network

The inter-datacentre network is used to interconnect different cloud infrastructures. These infrastructures may be owned by the same or different CSPs.

The inter-datacenter network requirements include:

- The inter-datacentre network is recommended to support the scalability to match the demand level of public and private clouds.
- The inter-datacentre network is recommended to be resilient;
- The inter-datacentre network is recommended to deal with virtual machine network addresses overlapping;
- The inter-datacentre network is recommended to support different logical networks.

9 Requirements for storage resources

This clause provides requirements for storage resources.

NOTE – An example of a reference model for storage resources is provided in Appendix I.

9.1 Storage space

The storage space requirement includes:

- It is required to support dynamic storage space expansion.

9.2 Storage interface

The storage interface requirements include:

- The storage resources are required to support either block storage interfaces or file system interfaces.
- The storage resources are recommended to support object storage accessed via web service data path interfaces.
- The storage resources are recommended to support structured data-sharing access interfaces.
- The storage resources can optionally support multiple types of interfaces.

9.3 Storage management

The storage management requirements include:

- It is required to provide the capabilities for user authentication and authorization.
- It is required to provide management capabilities for storage resources.
- It is required to provide basic configuration capabilities, including storage domain configuration, file system namespace configuration, storage resources configuration and local file system configuration.
- It is recommended to provide performance monitoring and statistics (e.g., disk I/O speed, disk space usage, CPU utilization, memory utilization, job completion).
- It is recommended to support alert capabilities, e.g., for event and trouble reporting.
- It is recommended to provide replication, archive and retention capabilities.

9.4 Storage availability

The storage availability requirements include:

- It is required to monitor data loss or failure.
- It is recommended to provide data backup and data recovery.
- It is recommended to provide data verification capabilities.
- It is recommended to support access through legitimate channels without time constraints, as well as the geographical constraints.
- It is recommended to support data synchronization to keep data consistency.

9.5 Data de-duplication

The data de-duplication is a method of reducing storage usage by eliminating redundant data. The data de-duplication can save resources of storage space and network bandwidth to transfer data.

The data de-duplication requirement includes:

- It is recommended for storage resources to support the data de-duplication capability.

10 Requirements for resources abstraction and control

Resources abstraction and control allows a CSP to access physical resources through software abstraction. It also provides composition, coordination, monitoring and scheduling of processing, storage and networking resources.

Resources abstraction and control directs the creation, modification, customization and release of abstracted resources. Resource abstraction and control is also responsible for controlling the interactions between resource pools and cloud services. A resource template refers to a set of standardized formatted hardware and software configuration settings for processing, storage and networking resources.

The resources abstraction and control requirements include:

- It is recommended that abstracted resources can be accessed and provisioned in a unified manner.
- It is recommended that abstracted resources are discovered, used and released through unified interfaces.
- It is recommended that abstracted resources are deployed and provisioned based on pre-defined policies.
- It is required to provide life-cycle management of resource templates (e.g., resource template creation, publication, activation, revocation and deletion).

- A resource template can optionally be applied to a group of resources at the same time.
- It is required to support monitoring of all physical and virtual resources.
- It is recommended that resource monitoring is capable of detecting the failures of resources.

11 Support of emergency telecommunications

Under emergency telecommunications (ET) [b-ITU-T Y.2205], any emergency-related service is understood as requiring special handling relative to other services.

If any component in the cloud infrastructure is used to support an emergency telecommunications service (ETS), the requirements in [b-ITU-T Y.1271] are relevant.

12 Security considerations

It is recommended that the security requirements of [b-ITU-T Y.2201], [b-ITU-T Y.2701] and the applicable X, Y and M series of ITU-T security Recommendations be taken into consideration; this includes access control, authentication, data confidentiality, communications security, data integrity, availability and privacy.

Security aspects for consideration within cloud computing environment are addressed by security challenges for CSPs, as described in [ITU-T X.1601]. In particular, [ITU-T X.1601] analyses security threats and challenges and describes security capabilities that could mitigate these threats and meet security challenges.

Appendix I

Overview and reference model for storage in a cloud environment

(This appendix does not form an integral part of this Recommendation.)

Storage resources are used to store a huge amount of data. The traditional storage system utilizes a tightly-coupled symmetry reference model which aims to work out high performance computing problems and may fulfil cloud computing scalability requirements. The next generation system adopts a loosely-coupled asymmetry reference model which centralizes metadata and controls manipulation. This reference model is not suitable for high performance computing; however, this design is to solve large capacity storage needs based on cloud computing deployment.

The applications and data in cloud environments need be delivered and maintained reliably using a tightly-coupled architecture. Other applications (e.g., search engines, media streaming) may rely on loosely-coupled architecture.

I.1 Reference model for cloud storage

Cloud storage delivers virtualized storage on demand over a network based on cluster, grid and distributed file systems. When the key issue of operation and processing in cloud computing is the storage and management of large-scale data, a large number of storage equipment need to be deployed. Hence, cloud storage is a cloud computing system for data storage and management.

Figure I.1 depicts the cloud storage reference model.

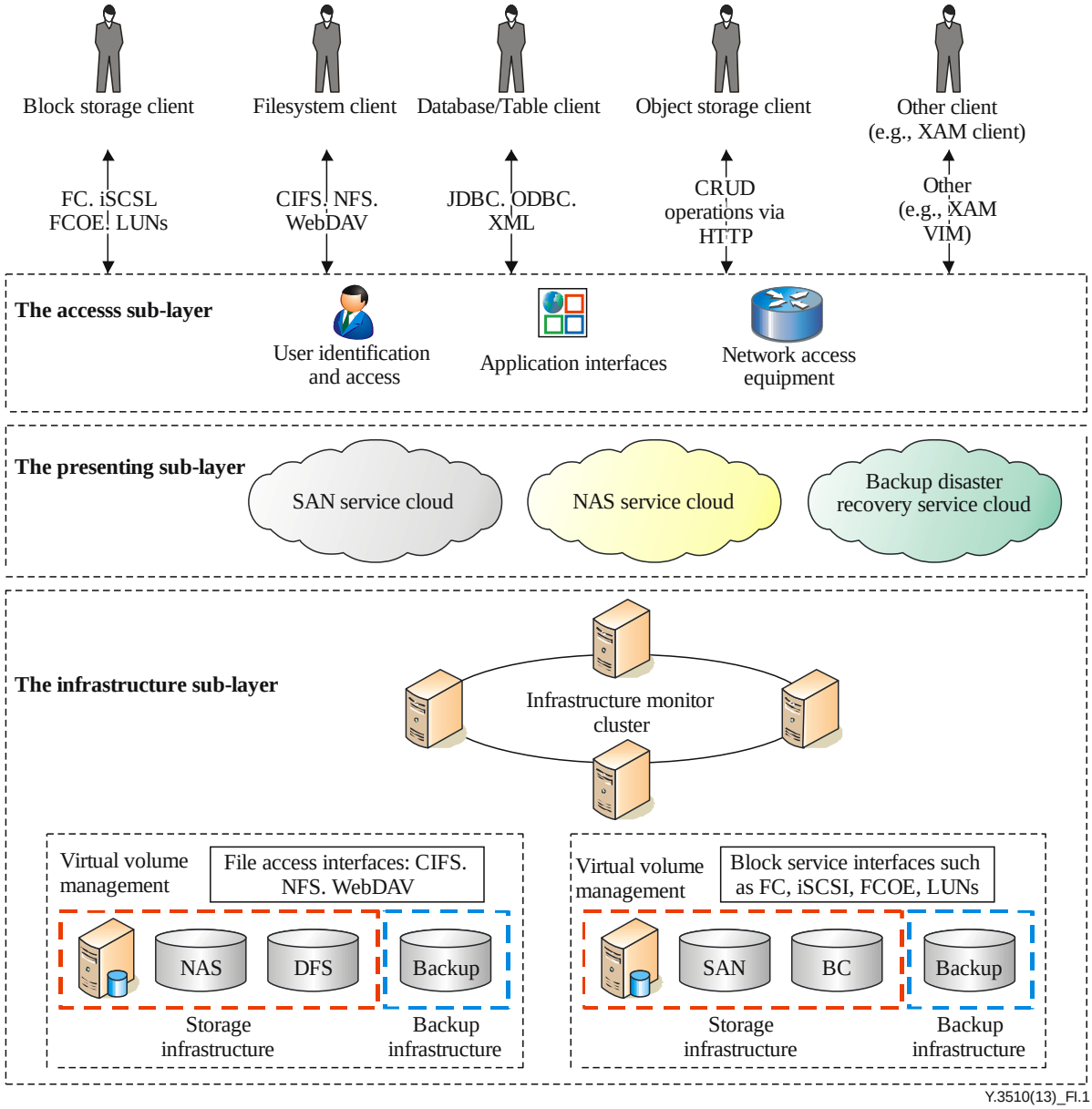

Figure I.1 – Cloud storage reference model

NOTE – The interfaces and protocols shown in Figure I.1 are examples used for illustrative purposes.

Cloud storage is the cooperating operation of multiple storage devices, multi-applications and multi-services. Not every storage system can be called cloud storage. A cloud storage system can provide functionalities such as a storage area network (SAN), network attached storage (NAS), data backup and disaster recovery.

As shown in Figure I.1, the cloud storage reference model is composed of three sub-layers which are described in the following clauses.

I.1.1 Infrastructure sub-layer

This sub-layer consists of the following 3 parts:

– The storage infrastructure is composed of common used storage devices such as fibre channel storage devices, NAS and Internet small computer system interface (iSCSI) [b-IETF RFC 3720] storage devices, as well as some related supporting appliances, such as switches for storage. Storage infrastructure will typically consist of several distributed working nodes to support high availability and reliability. A working node can include a virtual volume management element, an NAS and a distributed file system (DFS) device.

Another type of working node can include a virtual volume management element, a SAN and a block control device.

– The backup infrastructure is composed of a physical type library, a virtual type library, a database and related software.

– The infrastructure monitor cluster is composed of many servers which manage and monitor all kinds of storage and backup devices, repair related links and check the redundancy and carry out centralized management. It can include a global schedule function to provide the resources location in the storage infrastructure depending on the received accessing requests and associated requested resources. The servers typically support distributed hash table (DHT) networking to provide a general accessing interface for name space management, load balance, metadata management, routing management and duplication management. The infrastructure monitor cluster can access virtual volume management elements of the storage infrastructure to realize unified volume management and policy management.

I.1.2 Presenting sub-layer

This sub-layer is the core of the service logic of the cloud storage system. It provides several storage services, such as services based on SAN or NAS, as well as backup disaster recovery services.

SAN and NAS-based services provide key storage services for the management of cloud storage, detection and repair of the faulty links, status monitoring and QoS.

Backup disaster recovery services supply high-level data protection making unnecessary the use of a specialized disaster recovery network.

I.1.3 Access sub-layer

This sub-layer consists of storage-based application interfaces, network access equipment, user identification functions and other relevant access functions. Once authenticated and authorized, users make use of the cloud storage services, such as those based on the network file system (NFS) [b-IETF RFC 3530] or iSCSI [b-IETF RFC 3720].

The access sub-layer connects users to the presenting sub-layer through the use of private or public networks.

Appendix II

Considerations on resource monitoring

(This appendix does not form an integral part of this Recommendation.)

This appendix provides considerations on resource monitoring.

II.1 Health monitoring

Health monitoring of the cloud infrastructure includes monitoring the status of resources such as the physical server hardware, hypervisor, virtual machine, physical and virtual network switches and routers and storage systems.

A resource map displays all of the technology components, including transactions, applications, web servers, network switches, virtualized components and third-party cloud services. Having such a map can play an important role in effective business service management because when there is an application or transaction problem, it can help pinpoint the infrastructure components that may be playing a role in service disruptions.

In addition, the resource map is important to provide run-time monitoring, because cloud infrastructure is constantly changing. It is necessary to ensure the management of this resource map on a continuous basis. Non-intrusive probes can be used to automatically detect infrastructure, application and transaction changes in near real-time.

II.2 Performance monitoring

Basic performance monitoring looks at the CPU, memory, storage and network performance metrics from the VM guest OS, as well as from the hypervisor. These metrics typically get monitored even in non-virtualized environments. The virtualization-specific metrics could be for specific entities that are introduced by various virtualization technologies. The behaviour of other virtualization features can also be measured as metrics, such as how frequently VM migrations are occurring or when other availability features are engaged. Then there are specialized applications built by virtualization, for example, desktop virtualization. Monitoring for such solutions needs more parameters to be collected from the VM, as well as the hypervisor, for example, how quickly VMs are provisioned to a requesting end user.

II.3 Capacity monitoring

Resource utilization is continuously evolving. Therefore, the continuous planning of various resources such as servers, desktops, networks, storage and also many kinds of software is needed. This demands periodic audits of physical and virtual resources. Capacity monitoring needs end-to-end continuous capacity monitoring of the following key metrics:

- **Server utilization**: Peak and average server resource utilization, memory, CPU, resource, server bottlenecks and correlation with a number of VMs.
- **Memory usage**: Memory utilization on each server, capacity bottlenecks and relationship with a number of VMs and with different cloud services.
- **Network usage**: Peak and average network utilization, capacity/bandwidth bottlenecks and relationship with a number of VMs and with different cloud services.
- **Storage utilization**: Overall storage capacity metrics, VM and virtual disk utilization, I/O performance metrics, snapshot monitoring and correlation with a number of VMs and with different cloud services.

II.4 Security and compliance monitoring

Virtualization introduces a new set of security risks due to VM sprawl and the introduction of new threat targets such as the hypervisor layer, virtual infrastructure (VI) configurations and potential conflicts in the way access control is managed and policies are applied. Security and compliance monitoring becomes critical for securing the virtualized environment. Security and compliance monitoring needs end-to-end VI activity monitoring for:

- **VM sprawl**: Metrics to monitor VM activities as they get cloned, copied and, due to network migration, transfers to different storage media.
- **Configuration metrics**: Virtual server configuration monitoring to ensure that they are compliant with standards and hardening guidelines, VM configuration monitoring for software licensing policy enforcement and VI events that help enforce and detect violations of policy. This includes individual security and organization security policy monitoring.
- **Access control**: Access control monitoring and reports for role-based access control enforcement.
- **Compliance monitoring**: Metrics to validate audit and certification.

II.5 Monitoring and metering for charging and billing

In a virtualized environment the infrastructure is centralized and it is important to measure resource usage by different CSCs. This information can be used to distribute, amortize and in some cases, recover the cost correctly across the organization through a proper chargeback mechanism. Chargeback could be based on dynamic parameters such as resource usage and/or fixed parameters. To compute the correct chargeback information in a dynamic virtualized environment, it is important to monitor virtual and physical resource usage and allocations, as well as to be able to normalize the measurement across the cloud infrastructure. The monitoring and metering data for service charging should be collected and kept according to SLA objectives.

Chargeback monitoring needs end-to-end VI activity monitoring and service usage metering for:

- **Standard metrics**: All chargeable resource metrics like CPU usage, memory usage, storage usage and network usage metrics.
- **Key VI events**: VI events for virtual resource life-cycle events like start date and end date of VM creation and allocation.
- **Configuration monitoring**: VM configuration in terms of assigned resources and reservations and also applications installed to account for software licensing costs.
- **VM usage metrics**: VM uptime, number of VMs can vary depending on how the charging model is employed in the organization.

II.6 Monitoring in support of cloud services

The need for application and service monitoring is important in the cloud computing environment, especially for SLA/QoS evaluation because the application or service may have problems even if the VM or the physical server on which it is running looks normal. Application and service needs to monitor the basic health of application servers with the help of application-specific response time and throughput metrics. The analytics on this data could be used to correlate the application-observed and service-observed metrics to all layers of the infrastructure to perform a root-cause analysis in the event something going wrong. Application and service performance monitoring using the capture of network traffic is used more and more commonly in this area.

There are a few other aspects to virtual infrastructure monitoring that add to the complexity of building a comprehensive monitoring solution. All kinds of virtualization software allow the API to be able to collect metrics. However, each kind of virtualization software has its own object models. There are wide differences in features and even the behaviour of the common features. Therefore, the analytics that are to be built on the collected metrics must be developed for each kind of virtualization software.

Appendix III

Power management in cloud infrastructure

(This appendix does not form an integral part of this Recommendation.)

Datacentres are amongst the highest consumers of electricity all over the world. One distinct advantage of cloud computing is it is also able to power-manage hardware and devices. Therefore, the resources in a cloud infrastructure are recommended to be dynamically power-managed. The resources in cloud infrastructure are often arranged in trees. As some resources of a cloud infrastructure become idle, it can decrease power twigs or branches on trees. As the resource usage trends of cloud infrastructure are measured and controlled, it would be possible for these networks to put energy back into the grid by providing the grid with accurate time-based predictions of energy use. The grid can use this information to redirect energy to other destinations, or make other intelligent decisions.

Power management in cloud infrastructure represents a collection of processes and supporting technologies geared towards optimizing datacentre performance against cost and structural constraints. This includes increasing the deployable number of servers per rack when racks are subject to power or thermal limitations and making power consumption more predictable and easier to plan for.

Power management in cloud infrastructure comes in two categories: static and dynamic. Static power management deals with fixed power caps to manage aggregate power, while policies under dynamic power management take advantage of additional degrees of freedom inherent in virtualized cloud datacentres, as well as the dynamic behaviours supported by advanced platform power management technologies.

Appendix IV

Considerations on supporting of ETS

(This appendix does not form an integral part of this Recommendation.)

[b-ITU-T Y.1271] specifies the network requirements and capabilities to support ETS over both circuit-switched and packet-switched networks. Annex A of [b-ITU-T Y.1271] contains the list of functional requirements and categorizes them as essential and optional. Support for these requirements is needed for the scenario of when an ETS is offered by the CSP.

The requirements in [b-ITU-T Y.1271] can be separated into those relevant to the networking resources and those relevant to core transport network(s). Some requirements are applicable both for resources and for transport network(s). This clause considers the requirements for the networking resources category using the general requirements in clauses 8.1 and 8.2. The requirements relevant for networking resources from [b-ITU-T Y.1271] include: enhanced priority treatment, location confidentiality, restorability, interoperability, survivability/endurability, scalable bandwidth, reliability/availability and preferential treatment in congestion control measurement.

A cloud supporting ETS needs to be robust and able to support customers despite widespread damage. Another requirement is the restoration of access to cloud infrastructure resources including links connecting to the cloud. The processing nodes (virtual or physical) are to be restored quickly if damage to infrastructure resources occurs.

Cloud infrastructure resources need to adapt quickly for emergency applications, an adaptation that equates to application acceleration as noted in clause 8.1. Because ETS have requirements for different policies (QoS, security, traffic), the migration requirements in clause 8.1.5 are necessary to guarantee the SLAs among ETS customers and their CSPs.

The requirements, from [b-ITU-T Y.1271], specifically relevant to the support of ETS in core networks include: secure networks, restorability, network connectivity, mobility, coverage, survivability (connections), voice and data transmission, scalable bandwidth and reliability. Some of these requirements are applicable to both networking resources and the core transport network.

The requirements of clause 8.1, in relation to ETS, apply to ubiquitous coverage and therefore have the potential to preclude the need to establish special facilities after the occurrence of an emergency or disaster.

Reliability considerations of clause 8.1, in relation to ETS, are necessary for the network infrastructure to support survivability and endurability.

In support of ETS, the network should be smart enough for high priority applications. Some aspects of cloud services may be applicable to the offerings of priority services to facilitate disaster recovery functions, such as locating survivors and providing vital situational awareness information to government first responders and relatives of survivors affected by a disaster. Cloud computing can support complex modelling, analysis and rendering images to the first responders of disasters. [b-Tohoku]

Rapid authentication of authorized users for ETS implies awareness of the user/terminal attributes (subscriber profile data) and at the same time prevents unauthorized access, denial of service attacks and protection from intrusion.

Bibliography

[b-ITU-T E.107] Recommendation ITU-T E.107 (2007), *Emergency Telecommunications Service (ETS) and interconnection framework for national implementations of ETS*.

[b-ITU-T M.60] Recommendation ITU-T M.60 (1993), *Maintenance terminology and definitions*.

[b-ITU-T Q.1741.7] Recommendation ITU-T Q.1741.7 (2011), *IMT-2000 references to Release 9 of GSM-evolved UMTS core network*.

[b-ITU-T Y.1271] Recommendation ITU-T Y.1271 (2014), *Framework(s) on network requirements and capabilities to support emergency telecommunications over evolving circuit-switched and packet-switched networks*.

[b-ITU-T Y.2201] Recommendation ITU-T Y.2201 (2009), *Requirements and capabilities for ITU-T NGN*.

[b-ITU-T Y.2205] Recommendation ITU-T Y.2205 (2011), *Next Generation Networks – Emergency telecommunications – Technical considerations*.

[b-ITU-T Y.2701] Recommendation ITU-T Y.2701 (2007), *Security requirements for NGN release 1*.

[b-ITU-T Y.3011] Recommendation ITU-T Y.3011 (2012), *Framework of network virtualization for future networks*.

[b-IETF RFC 3530] IETF RFC 3530 (2003), *Network File System (NFS) version 4 Protocol*.

[b-IETF RFC 3720] IETF RFC 3270 (2004), *Internet Small Computer Systems Interface (iSCSI)*.

[b-IETF RFC 5905] IETF RFC 5905 (2010), *Network Time Protocol Version 4: Protocol and Algorithms Specification*.

[b-Tohoku] ACCJ (2011), *Responding to the Greater Tohoku Disaster: The Role of the Internet and Cloud Computing in Economic Recovery and Renewal. ACCJ Internet Economy Task Force*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Terminals and subjective and objective assessment methods
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities**
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2016