International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.3600

(11/2015)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

Cloud Computing

Big data – Cloud computing based requirements and capabilities

Recommendation ITU-T Y.3600

International Telecommunication Union

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

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**

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

Recommendation ITU-T Y.3600

Big data – Cloud computing based requirements and capabilities

Summary

Recommendation ITU-T Y.3600 provides requirements, capabilities and use cases of cloud computing based big data as well as its system context. Cloud computing based big data provides the capabilities to collect, store, analyse, visualize and manage varieties of large volume datasets, which cannot be rapidly transferred and analysed using traditional technologies.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y.3600	2015-11-06	13	11.1002/1000/12584

Keywords

Big data, big data ecosystem, cloud computing, data analytics, data storage, real-time analysis.

* 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 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.3600

Big data – Cloud computing based requirements and capabilities

1 Scope

This Recommendation provides an approach to use cloud computing to meet existing challenges in the use of big data. This Recommendation addresses the following subjects:

- Overview of big data;
  - Introduction to big data;
  - Big data ecosystem and roles;
  - Relationship between cloud computing and big data;
- Cloud computing based big data system context and benefits;
- Cloud computing based big data requirements;
- Cloud computing based big data capabilities.

Note that use cases of cloud computing based big data are provided in Appendix I and II.

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.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 activity [ITU-T Y.3502]: A specified pursuit or set of tasks.

3.1.2 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.

3.1.3 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 partner [ITU-T Y.3500]: Party which is engaged in support of, or auxiliary to, activities of either the cloud service provider or the cloud service customer, or both.

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

3.1.6 metadata [b-ITU-T M.3030]: Data that describes other data.

3.1.7 role [ITU-T Y.3502]: A set of activities that serves a common purpose.

3.1.8 sub-role [ITU-T Y.3502]: A subset of the activities of a given role.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 big data: A paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics.

NOTE – Examples of datasets characteristics include high-volume, high-velocity, high-variety, etc.

3.2.2 big data as a service (BDaaS): A cloud service category in which the capabilities provided to the cloud service customer are the ability to collect, store, analyse, visualize and manage data using big data.

3.2.3 party: Natural person or legal person, whether or not incorporated, or a group of either.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

API Application Programming Interface

BDaaS Big Data as a Service

BDAP Big Data Application Provider

BDC Big Data service Customer

BDIP Big Data Infrastructure Provider

BDSP Big Data Service Provider

BDSU Big Data Service User

CCRA Cloud Computing Reference Architecture

CDR Charging Detailed Record

CGF Charging Gateway Function

CSC Cloud Service Customer

CSN Cloud Service partner

CSP Cloud Service Provider

DP Data Provider

DPI Deep Packet Inspection

HTML Hyper Text Mark-up Language

IaaS Infrastructure as a Service

IoT Internet of Things

PDA Personal Digital Assistant

PDSN Packet Data Serving Node

SNS Social Network Service

SQL Structured Query Language

XML Extensible Markup Language

5 Conventions

The keywords "**is required to**" 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.

In the body of this document and its annexes, the words shall, shall not, should, and may sometimes appear, in which case they are to be interpreted, respectively, as is required to, is prohibited from, is recommended, and can optionally. The appearance of such phrases or keywords in an appendix or in material explicitly marked as informative are to be interpreted as having no normative intent.

6 Overview of big data

6.1 Introduction to big data

With the rapid development of information and communications technology (ICT), Internet technologies and services, huge amounts of data are generated, transmitted and stored at an explosive rate of growth. Data are generated by many sources and not only by sensors, cameras or network devices, but also by web pages, email systems and social networks as well as by many other sources. Datasets are becoming so large and so complex or are arriving so fast that traditional data processing methods and tools are inadequate. Efficient analytics of data within tolerable elapsed times becomes very challenging. The paradigm being developed to resolve the above issues is called big data.

For the purpose of this Recommendation it is understood, that within the big data ecosystem, data types include structured, semi-structured and unstructured data. Structured data are often stored in databases which may be organized in different models, such as relational models, document models, key-value models, graph models, etc. Semi-structured data does not conform to the formal structure of data models, but contain tags or markers to identify data. Unstructured data do not have a pre-defined data model and are not organized in any defined manner. Within all data types data can exist in formats, such as text, spreadsheet, video, audio, image, map, etc.

Big data are successfully used in many fields, if traditional methods and tools have become inefficient, where data processing is characterized by scale (volume), diversity (variety), high speed (velocity) and possibly other criteria like credibility (veracity) or business value. These characteristics, usually called the Vs, can be explained as follows:

– Volume: refers to the amount of data collected, stored, analysed and visualized, which big data technologies need to resolve;

– Variety: refers to different data types and data formats that are processed by big data technologies;

– Velocity: refers to both how fast the data is being collected and how fast the data is processed by big data technologies to deliver expected results.

NOTE – Additionally, veracity refers to the uncertainty of the data and value refers to the business results from the gains in new information using big data technologies. Other Vs can be considered as well.

Taking into account the above Vs' described characteristics, big data technologies and services allow many new challenges to be resolved and also create more new opportunities than ever before:

– Heterogeneity and incompleteness: Data processed using big data can miss some attributes or introduce noise in data transmission. Even after data cleaning and error correction, some incompleteness and some errors in data are likely to remain. These challenges can be managed during data analysis. [b-CRA-BDWP]

– Scale: Processing of large and rapidly increasing volumes of data is a challenging task. Using data processing technologies, the data scale challenge was mitigated by the evolution of processing and storage resources. Nowadays however data volumes are scaling faster than resources can evolve. Technologies such as parallel databases, in memory databases, non-SQL databases and analytical algorithms allow this challenge to be resolved.

– Timeliness: The acquisition rate and timeliness, to effectively find elements in limited time that meet a specified criterion in a large dataset, are new challenges faced by data processing. Other new challenges are related to the types of criteria specified and there is a need to devise new index structures and responses to the queries having tight response time limits.

– Privacy: Data about human individuals, such as demographic information, Internet activities, commutation patterns, social interactions, energy or water consumption, are being collected and analysed for different purposes. Big data technologies and services are challenged to protect personal identities and sensitive attributes of data throughout the whole data processing cycle while respecting applicable data retention policy.

Positive resolving of the above challenges opens new opportunities to discover new data relationships, hidden patterns or unknown dependencies.

6.2 Big data ecosystem

This clause describes an environment, called the big data ecosystem through roles and sub-roles. It also defines necessary activities for roles providing and consuming big data services as well as relationships between roles.

The big data ecosystem includes the following roles:

– data provider;

– big data service provider;

– big data service customer.

The big data ecosystem is shown in Figure 6-1.

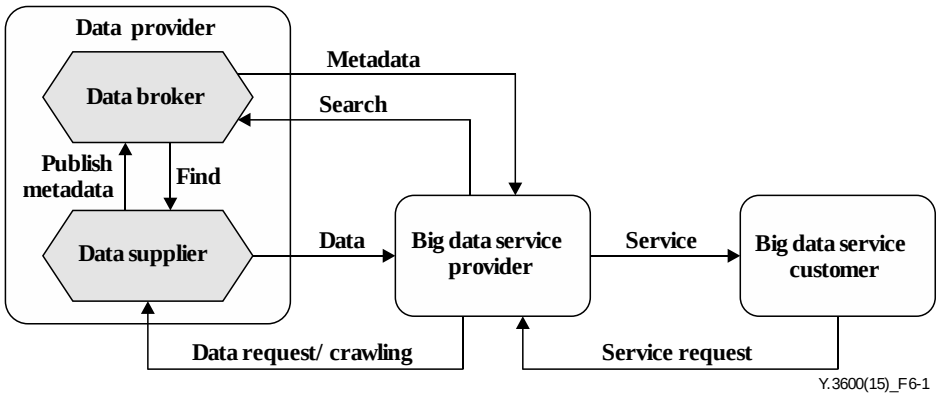


Figure 6-1 – Big data ecosystem

6.2.1 Data provider (DP)

The data provider (DP) role consists of two sub-roles:

– data supplier;

– data broker.

6.2.1.1 Data supplier

The data supplier provides data from different sources to the data broker, which can be accessed by the big data service provider. The data supplier's activities include:

– generating data;

– creating metadata information describing the data source(s) and relevant attributes;

– publishing metadata information to access the metadata.

6.2.1.2 Data broker

The data broker serves to connect between the data supplier and the big data service provider. The data broker can act as a clearinghouse, open data mart, etc. and its activities include:

– providing a meta-information registry to the data supplier for publishing their data sources;

– finding on-line open-data sources and registering corresponding meta-information;

– providing a service catalogue to the big data service provider for searching usable data.

6.2.2 Big data service provider (BDSP)

The big data service provider (BDSP) supports capabilities for big data analytics and infrastructure. The big data service provider can act as a form of big data platform, an extension of the existing data analytics platform, etc. Big data service provider activities include:

– searching data sources (from the data broker) and collecting data by requesting and crawling;

– storing data to a data repository;

– integrating data;

– providing tools for data analysis and visualization;

– supporting data management such as data provenance, data privacy, data security, data retention policy, data ownership, etc.

6.2.3 Big data service customer (BDC)

The big data service customer (BDC) is the end-user or is a system that uses the results or services from a big data service provider. The big data service customer may produce new services or

knowledge on consumer activities and furnish them outside of the big data ecosystem. Big data service customer activities include:

– requesting big data services from the big data service provider;

– using the outputs of big data services.

6.3 Relationship between cloud computing and big data

Big data refers to technologies and services which extract valuable information from the extensive datasets characterized by the Vs, while cloud computing is, as defined in [ITU-T Y.3500], the 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.

Big data needs on-demand high performance data processing and distributed storage as well as variety of tools required to accomplish activities of the big data ecosystem which are described in clause 6.2. Cloud computing meets the challenges of big data as described in clause 6.1. The burst nature of workloads makes cloud computing more appropriate for big data challenges such as scalability and timeliness. The big data ecosystems, which are supported by a cloud computing system context, can be referred to as cloud computing based big data. Cloud computing based big data is addressed in detail in clause 7.

7 Cloud computing based big data

This clause describes a cloud computing based big data system context that is effective for supporting big data. It also provides benefits of cloud computing based big data.

7.1 Cloud computing based big data system context

Cloud computing based big data system context is described with new sub-roles and activities based on the architectural user view defined in [ITU-T Y.3502]. This clause describes how cloud computing can support the three main roles in a big data ecosystem: data provider, big data service provider and big data service customer.

Cloud computing sub-roles can be mapped to big data roles as shown in Table 7-1.

Table 7-1 – Mapping table between big data ecosystem roles and cloud computing based big data system context sub-roles

Big data ecosystem roles	Cloud computing based big data system context sub-roles
Data provider	CSN:data provider
Big data service provider	CSP:big data infrastructure provider, CSP:big data application provider
Big data service customer	CSC:big data service user

Figure 7-1 illustrates the cloud computing sub-roles for big data. Figure 7-1 also identifies activities specific for big data and assigns them to cloud computing sub-roles. Arrows used in Figure 7-1 show the data and service flows within a cloud computing based big data system context.

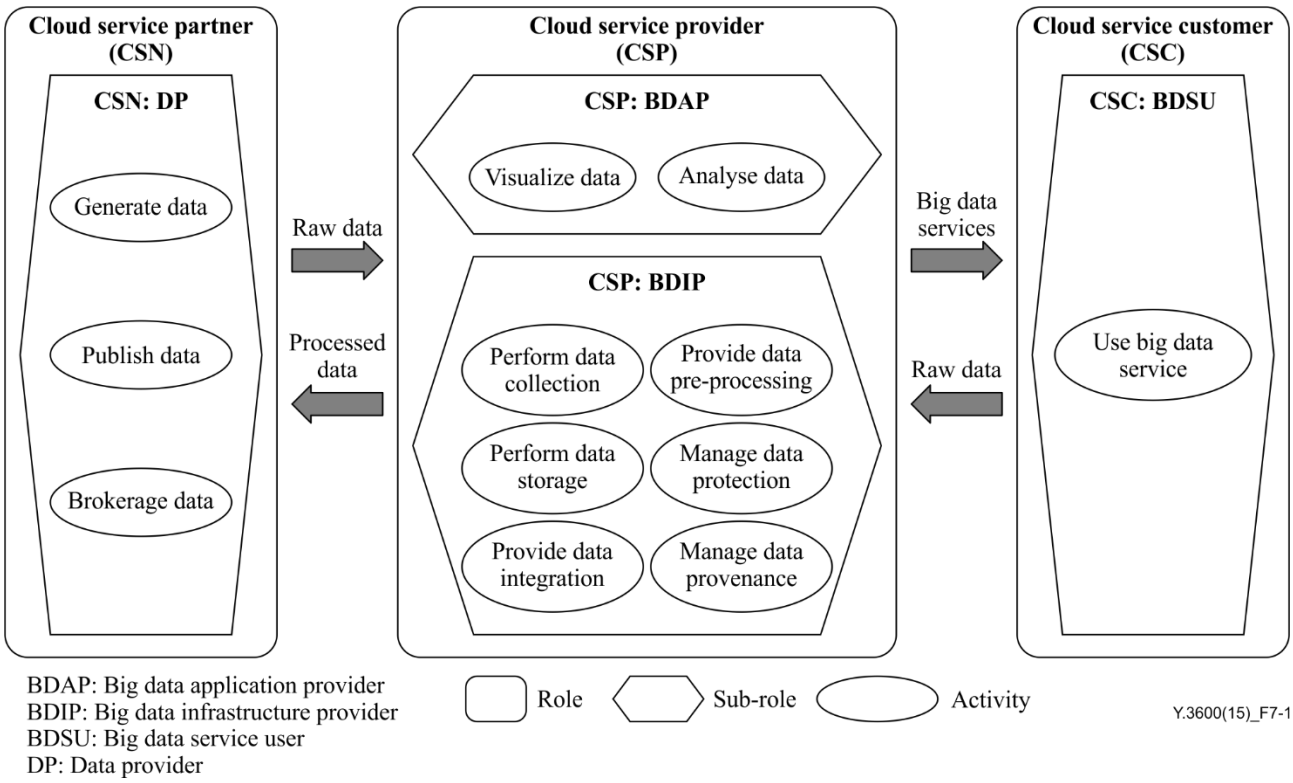


Figure 7-1 – Cloud computing based big data system context

Figure 7-2 illustrates, as an example scenario, how cloud computing supports big data services from the viewpoints of CSP: Big data application providers (BDAPs) and CSP:Big data infrastructure providers (BDIPs). A cloud based big data service, which is provided by a CSP:BDAP or a CSP:BDIP, utilizes other sub-roles of the cloud service provider.

The CSP provides services of application and platform capability types such as software as a service and platform as a service and these services can be used by the CSP:BDAP to perform data analysis, visualization and other big data applications. In addition, the CSP:BDIP can use the cloud services of cloud infrastructure capability types such as compute as a service, data storage as a service, infrastructure as a service and network as a service to perform big data services for data collection, data processing, data management, etc.

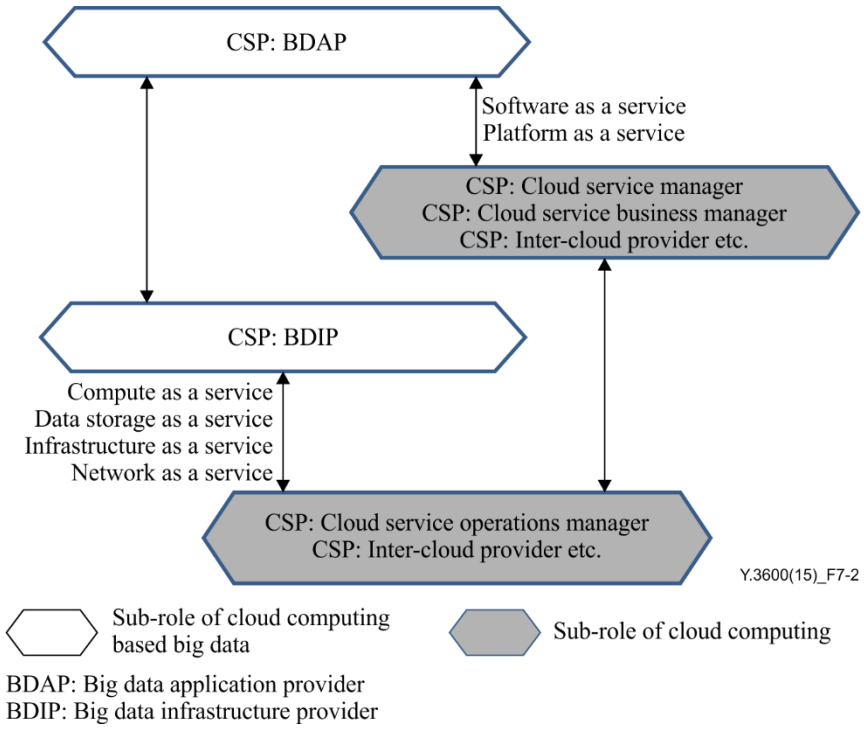


Figure 7-2 – An example scenario of cloud computing support for big data

NOTE – Figure 7-2 considers cloud service categories based on Table A.1 of ITU-T Y.3500.

7.1.1 CSN:data provider (CSN:DP)

CSN:data provider (CSN:DP) is the sub-role of the cloud service partner (CSN) generating and publishing new data or information which feeds into the big data system for discovery, access and transformation. The CSP:data provider's activities include a generate data activity, a publish data activity and a brokerage data activity.

7.1.1.1 Generate data activity

The generate data activity involves gathering data from several kinds of sources. The data can be generated in a variety of types such as structured data, semi-structured data and unstructured data.

The data sources include public data (data from governments, organizations, Internet, etc.), private/enterprise data and application based data, such as social network service (SNS) and Internet of Things (IoT) data.

7.1.1.2 Publish data activity

The publish data activity is the process of registering metadata of data to the CSN. It provides metadata for brokerage data activity.

NOTE – metadata is delivered to the CSP:BDIP through brokerage data (see clause 7.1.1.3) with the data catalogue which includes data access methods, data use policy, etc.

7.1.1.3 Brokerage data activity

The brokerage data activity involves providing metadata of data to the CSP: BDIP.

Types of brokerage data activities include:

– providing a data registry to the CSN:DP for publishing their data sources;
– finding an on-line data source and registering its metadata;
– providing a catalogue to the CSP:BDIP for searching appropriate data.

7.1.2 CSP: big data application provider (CSP:BDAP)

The CSP: big data application provider (CSP:BDAP) is a sub-role of the CSP executing a set of data lifecycle operations to meet the requirements of data analysis and visualization, as well as the security and privacy requirements. The CSP:BDAP utilizes the services from the CSP:BDIP for big data services and provides analysis result to the CSC: big data service user (CSC:BDSU).

Types of CSP:BDAP activities include:

– visualizing data;

– analysing data.

7.1.2.1 Visualize data activity

The visualize data activity is responsible for presenting data with multiple styles, such as statistical graphics, forms, diagrams, charts, reports, etc., through visualization and reporting tools and services. The big data service user (BDSU) can easily and quickly understand the meaning of data through data visualization. Direct interaction between reporting tools and CSC operational systems is also supported.

Types of big data visualize activities include:

– business intelligence;

– data reporting;

– data exploration.

7.1.2.2 Analyse data activity

The analyse data activity is a process to extract and discover useful information or valuable insights from big data.

Types of big data analyse activities include:

– statistical analysis;

– predictive analysis;

– content analysis.

7.1.3 CSP:big data infrastructure provider (CSP:BDIP)

CSP:big data infrastructure provider (CSP:BDIP) is the sub-role of the CSP providing cloud computing resources (such as system hardware, a network, storage and computing platforms) in order to execute big data processing, while protecting the privacy, retention policy and integrity of data.

Types of CSP:BDIP activities include:

– performing data collection;

– performing data storage;

– providing data pre-processing;

– providing data integration;

– managing data protection;

– managing data provenance.

7.1.3.1 Perform data collection activity

The perform data collection activity is responsible for gathering and searching of data. Data in a cloud computing environment that could be transported through Internet, such as web, video, audio, radio frequency identification (RFID), sensor and social network data and data obtained in other ways can be categorized into three types as follows:

- structured data, such as record data persistent in databases;
- semi-structured data, such as data stored in XML, HTML and other format documents;
- unstructured mass data, such as log files, video and audio data.

7.1.3.2 Perform data storage activity

The perform data storage activity allows storing of the data collected and the pre-processing results using storage resources and building of corresponding databases to manage and manipulate the data.

7.1.3.3 Provide data pre-processing activity

The provide data pre-processing activity realizes extraction, transformation and de-noising of the collected data. Since there are many different data formats and data types of the collected data, the transformation and extraction of complex data into simple structured data facilitates and speeds up the data analysis process.

The data de-noising process filters out the defects in data to eliminate negative effects in the normal analysis process.

7.1.3.4 Provide data integration activity

The provide data integration activity is responsible for combining, forming, coordinating or blending data from disparate sources and for solving the issues of bulk data movement, replication, synchronization, virtualization, data quality and data services.

7.1.3.5 Manage data protection activity

The manage data protection activity is responsible for protecting data so that the protected data may not be collected, stored and disclosed to whom may not be appropriate.

7.1.3.6 Manage data provenance activity

The manage data provenance activity manages information about the origin and generation process methods of data. Such information is useful for debugging, transformations, auditing, evaluating the quality of and trust in data, modelling authenticity and implementing access control for derived data.

7.1.4 CSC:big data service user (CSC:BDSU)

CSC:big data service user (CSC:BDSU) is the sub-role of the CSC performed by end-users or other systems in order to use the services from the CSP:BDAP and CSP:BDIP.

Types of CSC:BDSU activities include:

– use big data service;

7.1.4.1 Use big data service

The use big data service involves using the services of a CSP:BDAP and a CSP:BDIP in order to accomplish tasks.

Use big data service activity typically involves:

- provision of user credentials to enable the CSP:BDAP and CSP:BDIP to authenticate the user and grant access to the big data service;
- invocation of the big data service.

7.2 Benefits of cloud computing based big data

One of the main purposes of big data is to extract deep information from large volumes of data. Large volumes of data can be analysed using two key methods; the scale-up method and the scale-out method. The scale-up method uses a large system with enough resources to analyse a huge amount of data. In contrast, the scale-out method uses many separate processing nodes, where each

processing node analyses a portion of data. The scale-out method has the ability to scale just by adding more processing nodes. Cloud computing can provide big data with cost-effective elastic processing, storage and network resources.

The key benefits of cloud computing based big data are:

– **Scalability**. Big data needs to have capabilities to store and process large volumes of data. Therefore, scalability is very important for big data. However, the additional systems for big data require a lot of time and cost management. Cloud computing can provide flexible scalabilities to big data without additional expansion of infrastructure. It allows the big data service user to easily upscale or downscale the resources quickly.

– **Resiliency**. Cloud computing can support big data to have resiliency capabilities to maintain an acceptable level of service in the face of faults affecting normal operation.

– **Cost effectiveness**. Big data facilitates fast and scalable data processing such as system log analysis and click streams analysis. For many systems and platforms, there are huge volumes of log data and traditionally databases are used to perform log analysis. But the cost to perform data analysis (including costs of storage, system maintenance, etc.) is too high when traditional mechanisms are used. Cloud computing can offer flexible and scalable resources in a cost effective manner.

– **Efficient analysis**. In order to extract more valuable insights, big data applications and services need a well-defined analytic strategy as well as processing power. The cloud computing based big data service may dynamically use the required resources.

– **Deep information extraction**. Big data develops new business insights and mechanisms including prediction and decision assistance. This is different from conventional systems because the data processing logic to handle the raw data and what kind of information can be extracted from datasets is already known.

8 Requirements of cloud computing based big data

8.1 Data collection requirements

The data collection requirements include:

1) It is required for the CSP:BDIP to support collecting data from multiple CSN:DPs in parallel;

2) It is recommended for the CSN:DP to expose data to the CSP:BDAP by publishing metadata;

3) It is recommended that the CSP:BDIP supports collecting data from different CSN:DPs with different modes;

   NOTE – Data could be collected in different modes, such as pull mode in which the data collection process is initiated by CSP:BDIP, or push mode in which the data collection process is initiated by the CSN:DP.

4) It is recommended for the CSN:DP to provide a brokerage service to the CSP:BDIP for searching accessible data;

   NOTE – Brokerage provides data a catalog which has data information such as data specification, data instructions, electronic access methods, license policy, data quality, etc.

5) It is recommended that the CSP:BDIP integrates data delivered by the CSC and data publicly available;

6) Data collection can optionally be performed by the CSP:BDIP in real-time.

8.2 Data pre-processing requirements

The data pre-processing requirements include:

1) It is required for the CSP:BDIP to support data aggregation;

NOTE – Data from different sources can be organized in the same model or data format, as described in clause 6.1.

2) It is recommended that the CSP:BDIP provides the dedicated resources for pre-processing;

NOTE – Pre-processing includes extraction, transformation and de-noising of the collected data.

3) It is recommended that the CSP:BDIP supports unification of data collected in different formats;

NOTE –Unification of data is for example to unify data about persons/locations/dates extracted from web pages, pictures, videos, SNS data and calling logs to text format.

4) It is recommended for the CSP:BDIP to support extraction of data from unstructured data or semi-structured data into structured data.

NOTE – This requirement can be applied also to data storage.

8.3 Data storage requirements

The data storage requirements include:

1) It is required for the CSP:BDIP to support different data types with sufficient storage space, elastic storage capacity and efficient control methods;

2) It is required for the CSP:BDIP to support storage for different data formats and data models;

NOTE – Data formats include text, spreadsheet, video, audio, image, map, etc. Data models include relational models, document models, key-value models, graph models, etc. (as described in clause 6.1).

3) It is required that the CSP:BDIP provides a flexible licensing policy for the databases;

NOTE – As database systems may be covered by vendor licenses, the CSP:BDIP that offers a database as part of the big data service needs the ability to adapt the licensing conditions to the particular service and the CSC:BDSU requirements.

4) It is recommended that the CSP:BDIP provides different types of databases;

NOTE – Examples of database types include relational databases (RDB), object relational databases (ORDB), object oriented databases (OODB), NoSql (not only SQL) databases, etc.

5) It is recommended for the CSN:DP to expose application programming interfaces (APIs) for data delivery;

6) It is recommended that the CSP:BDIP fulfils storage and database performance demands.

7) It is recommended that the CSP:BDIP supports a data retention policy covering a data retention period before its destruction after termination of a contract. This is to protect the big data service customer from losing private data through an accidental lapse of the contract.

8.4 Data analysis requirements

The data analysis requirements include:

1) It is required for the CSP:BDAP to support analysis of various data types and formats;

2) It is required for the CSP:BDAP to support batch processing;

3) It is required for the CSP:BDAP to support association analysis;

NOTE – Association analysis is the task of uncovering relationships among data.

4) It is required for the CSP:BDAP to support different data analysis algorithms;

NOTE – Data analysis algorithms include classification, clustering, regression, association, ranking, etc.

5) It is required that the CSP:BDAP provides flexible licensing policy for the analytical applications;

6) It is recommended for the CSP:BDAP to support user defined algorithms;

7) It is recommended for the CSP:BDAP to support data processing in distributed computing environments;

8) It is recommended for the CSP:BDAP to support data indexing;

9) It is recommended that the CSP:BDAP supports data classification in parallel;

10) It is recommended that the CSP:BDAP provides different analytical applications;

11) It is recommended that the CSP:BDAP supports customization of analytical applications;

12) It is recommended for the CSP:BDAP to support real-time analysis of streaming data;

13) It is recommended for the CSP:BDAP to support user behaviour analysis;

NOTE – User behaviour includes user-related information, collected users' behaviour in real-time, environmental information and the analysed information from the cumulative users' information on a CSP:BDIP's storage. Scope of behaviour analysis is based on the user's agreement in advance.

14) The CSP:BDAP can optionally perform analysis of different data types and formats in real-time.

8.5 Data visualization requirements

The data visualization requirements include:

1) It is required that the CSP:BDAP provides a flexible licensing policy for the reporting tool;

2) It is recommended that the CSP:BDAP supports different tools or plug-ins with multiple styles of data visualization;

NOTE – Visualization styles include statistical graphics, forms, diagrams, charts, etc.

3) It is recommended that the CSP:BDAP supports customization of the reporting tools;

4) It is recommended that the CSP:BDAP supports integration of reporting tools with CSC reporting systems;

5) It is recommended that the CSP:BDAP supports integration of reporting tools with CSC operational systems;

6) It is recommended that the CSP:BDAP supports composed services which could combine two or more big data services to the CSC:BDSU.

8.6 Data management requirements

The data management requirements include:

1) It is required for the CSP:BDIP to manage metadata information such as creating, controlling, attributing, defining and updating;

NOTE – Metadata contains critical information such as persistent identification of the data, the fixity and the access rights of the stored data.

2) It is required for the CSP:BDIP to track a data history which contains the source of data and the data processing method;

3) It is required for the CSP:BDAP to support distributed cluster monitoring tools to monitor the health and status of computing clusters;

4) It is required for the CSP:BDIP to support data preservation policy management rules;

NOTE – Provided rules include data retirement and refreshment methods.

5) It is recommended for the CSP:BDIP to support network resource monitoring;

6) It is recommended for the CSP:BDIP to support management of data lifecycle operations;

NOTE – Data lifecycle operations include data generation, transmission, storage, use and deletion.

8.7 Data security and protection requirements

The data security and protection requirements include:

1) It is required for the CSP:BDIP to protect data collection, data storage, data transmission and data processing with security mechanisms;
2) It is required for the CSP:BDIP to support data protection;
3) It is required that the CSP deletes CSC related data and analytical results according to the lifetime defined by the CSC or on the CSC's demand;
4) It is recommended that the CSP supports implementing the CSC's data protection and security policies over data and analytical results;
5) It is recommended that the CSP:BDIP supports redundancy mechanisms and transaction logging.

9 Cloud computing based big data capabilities

9.1 Data collection capabilities

Data collection capabilities include:

– Data source intelligent recognition, which offers the capabilities to locate the data sources and detect the types of data being collected;
– Data adaptation, which offers the capabilities to transform and organize the data being collected with targeted data structures and attributes (numbering, location, ownerships, etc.);
– Data integration, which offers the capabilities to integrate data from different data sources (different data types) using metadata or ontology;
– Data brokerage, which offers the capabilities to provide a brokerage service for searching data.

9.2 Data pre-processing capabilities

Data pre-processing capabilities include:

– Data extraction, which offers the capabilities to extract information from the semi-structured data or unstructured data;
– Data transmission, which offers the capabilities to transport datasets (static data and real-time data) from data sources or between one location to another keeping the integrity and consistency;
– Data de-noising, which offers the capabilities to eliminate noise information from a mixture of signal data and noise data;
– Data aggregation, which offers the capabilities to aggregate data which come from different sources in the same data model or data format.

9.3 Data storage capabilities

Data storage capabilities include:

– Data storing, which offers the capabilities to store different types and formats of data with elastic storage capacity;
– Data registration, which offers the capabilities to create, update and delete the metadata with corresponding changes in data storage;

NOTE – In the case of unstructured data registration, the data registration component can request the transforming of raw data to semi-structured data such as JavaScript object notation (JSON) or binary JavaScript object notation (BSON) to define semantic relationships among different datasets for knowledge sharing.

– Data access, which offers the capabilities to access data through multiple interfaces, such as web service interfaces, file system interfaces, database interfaces and so on;
– Data indexing, which offers the capabilities to create and update indexes for datasets;
– Data duplication and backup, which offers the capabilities to duplicate and make backups for datasets.

9.4 Data analytics capabilities

Data analytics capabilities include:

– Data preparation, which offers the capabilities to transform data into a form that can be analysed. These capabilities include exploring, changing and shaping of the raw data;
– Data analysis, which offers the capabilities of investigation, inspection and modelling of data in order to discover useful information;
– Workflow automation, which offers the automation processes, in whole or part, during which data or functions are passed from one step to another for actions, according to a set of procedural rules;
– Analysis algorithm adaptation, which offers the capabilities to apply algorithms of classification, regression, clustering, association rules, ranking, etc. to process the datasets according to the CSC demands;
– Distributed processing, which offers the capabilities to distribute the processing tasks to a cluster of computing nodes;
– Data application, which offers the capabilities to support applications or application plug-ins to use the analysis results of datasets.

9.5 Data visualization capabilities

Data visualization capabilities include:

– Data visualization, which offers the capabilities to create, configure, deliver and customize the visual representation of data analysis results.
– Data reporting, which offers the capabilities to make reports of summary, key elements and analysis results of datasets.

9.6 Data management capabilities

Data management capabilities include:

– Data provenance, which offers the capabilities to manage information pertaining to any source of data including the party or parties involved in generation and introduction processes for data;
– Data preservation, which offers the capabilities to manage the series of activities necessary to ensure continued access to data according to relevant policy;
– Data ownership, which offers the capabilities to manage property rights of data possession and disposition according to the change of data status (e.g., after data integration);
– Processes monitoring, which offers the information related to data processing;

NOTE – This capability can include information such as the success of the job and task, running time, resource utilization, etc.

– Metadata management, which offers the capabilities of creating, defining, attributing, controlling and updating metadata information.

9.7 Data security and protection capabilities

Data security and protection capabilities include:

– Access control, which offers the capabilities to manage the rights of parties to control or influence the information related to them;

– Policy control, which offers the capabilities to control policies of data protection and security;

– Data security, which offers the capabilities to apply the storage, network and service related security mechanisms, including administrative, operational and maintenance issues.

10 Security considerations

Security aspects for consideration within cloud computing environments, including cloud infrastructure, IaaS, NaaS, DaaS 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.

Relevant security requirements of [b-ITU-T Y.2201], [b-ITU-T Y.2701] and applicable X, Y and M series of ITU-T Recommendations need to be taken into consideration, including access control, authentication, data confidentiality, data retention policy, network security, data integrity, availability and privacy.

Appendix I

Use cases of cloud computing in support of big data

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

Table I.1 – Personalization customized service using big data

Title	Personalization customized service using big data
Description	There are lots of network devices such as laptop computers, smart phones, smart pads, PDAs, health equipment, etc. Owners of these devices can access Internet and support network services in order to get some information, to buy something, to see movies, etc. Each activity might be logged such as network access information, records of visiting a website as well as usage of a service category in network service providers. This data volume can be of a tremendous size due to so many devices accessing Internet at same time. Consequently, data processing can become a big data problem. More meaningful information can however be extracted from big data. In this scenario, the CSP:BDIP should have capabilities to collect data and store said data, and the CSP:BDAP should have capabilities to analyse data which are related to the web service user's activities. The CSP:BDAP also has the task of providing the results of analysis to the CSC:BDSU. CSP:BDIP can store relevant information of the web service user's activities in the big data storage and provides this data to the CSP:BDAP. The CSP:BDAP can receive requests for operations with data and get some other information (e.g., weather information, season, holiday information, etc.) from external data sources which are supported by the CSN:DP. The CSP:BDAP may know the patterns or preferences of the user through analysis of the user's activities and pre-stored information. The CSP:BDAP can then provide the user's pattern information and the user's preferences to the CSC:BDSU. Such information is very helpful in enabling the CSC:BDSU to advertise to or support the web service users.
Roles/sub-roles	– CSP:BDIP – CSP:BDAP – CSN:DP – CSC:BDSU

Table I.1 – Personalization customized service using big data

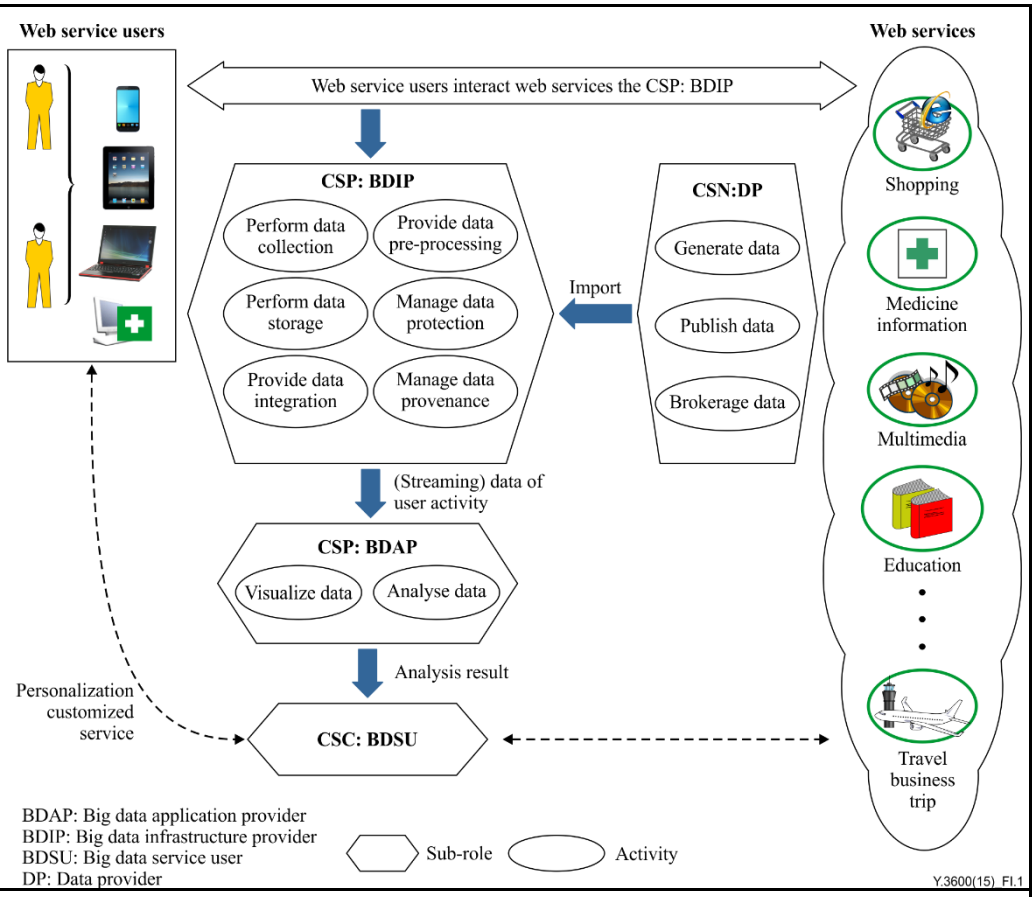
Figure	
Pre-conditions (optional)	– The CSP:BDIP can recognize the source and destination of the packet from the user under permission using personal user information
Post-conditions (optional)	
Derived requirements	– Data collection (refer to clause 8.1, requirements 1, 2) – Data preprocessing (refer to clause 8.2, requirements 1, 4) – Data storage (refer to clause 8.3, requirement 1, clause 8.7, requirement 3) – Data analysis (refer to clause 8.4, requirement 1, 12) – Data application (refer to clause 8.5, requirement 6, clause 8.1, requirement 4) – Data management (refer to clause 8.6, requirements 2, 4)

Table I.2 – The massive and high resolution multimedia data service

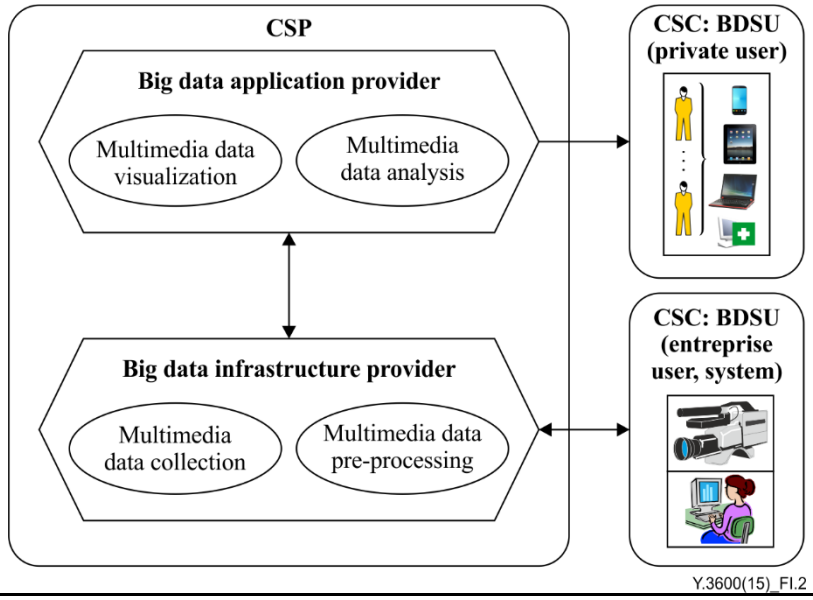
Title	The massive and high resolution multimedia data service
Description	In this use case, the cloud service provider should support the CSP:BDIP provider to collect, process and transmit the massive multimedia data using cloud services. In addition, the CSP should support the CSP:BDAP to provide massive multimedia related application services. The CSP:BDAP can provide the multimedia analysis and visualization to the user and provide the multimedia collection (transmission), processing and storage to the user or the CSP:BDAP.
Roles/sub-roles	– CSP – CSP:BDIP – CSP:BDAP – CSC:BDSU
Figure	
Pre-conditions (optional)	
Post-conditions (optional)	
Derived requirements	– Acceleration processing (refer to clause 8.2, requirement 2) – Network monitoring (refer to clause 8.6, requirement 5)

Table I.3 – E-commerce platform big data analysis

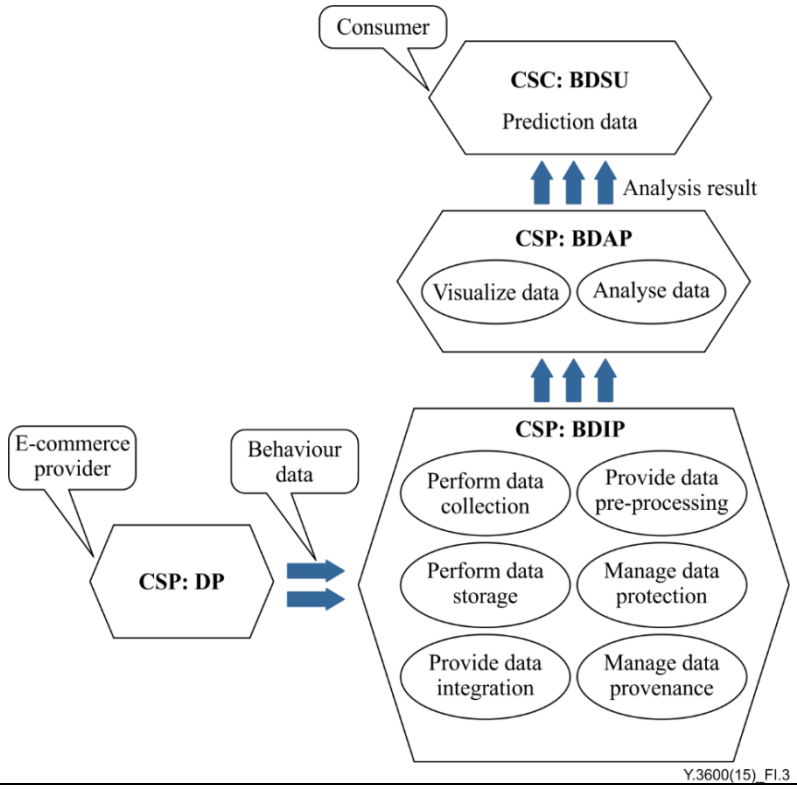
Title	E-commerce platform big data analysis
Description	Consumers shop online using an e-commerce platform, including search and checking of product information. The CSC:BDSU (e-commerce platform providers) use big data applications to implement backend supporting services, such as product recommendations, sales volume prediction, user behaviour analysis, etc. Customers get online shopping related services, such as authentication, products searching, checking, ordering and shipping through the front-end functions of the e-commerce platform. The e-commerce provider manages (creates, reads, updates and deletes) the information about products and inventory information through back-end functions of the e-commerce platform. Related big data application services are built upon the private or public big data infrastructure provided by the CSP:BDIP. The operational data of the e-commerce platform is collected, stored and pre-processed by the CSP:BDIP. In addition the CSP:BDAP analyses the operational data to give the prediction information. Privacy rules are followed, if applicable.
Roles/sub-roles	– CSN:DP – CSP:BDAP – CSP:BDIP – CSC:BDSU
Figure	
Pre-conditions	– The CSN:DP can provide user consumption behaviour data such as shopping frequency, often purchased items, payment capacity, etc.
Post-conditions	– The CSP:BDAP can have responsibility for providing analysis result of user consumption behaviour data such as prediction information and user behaviour analysis. – The CSC:BDSU gives accurate and personalized sales and promotions such as product recommendations, sales volume prediction and user behaviour analysis.
Derived requirements	– Data collection (refer to clause 8.1, requirements, 3, 6) – Data analysis (refer to clause 8.4, requirements 8, 9, 12) – Data management (refer to clause 8.7, requirement 5)

Table I.4 – Internet social network services big data analysis

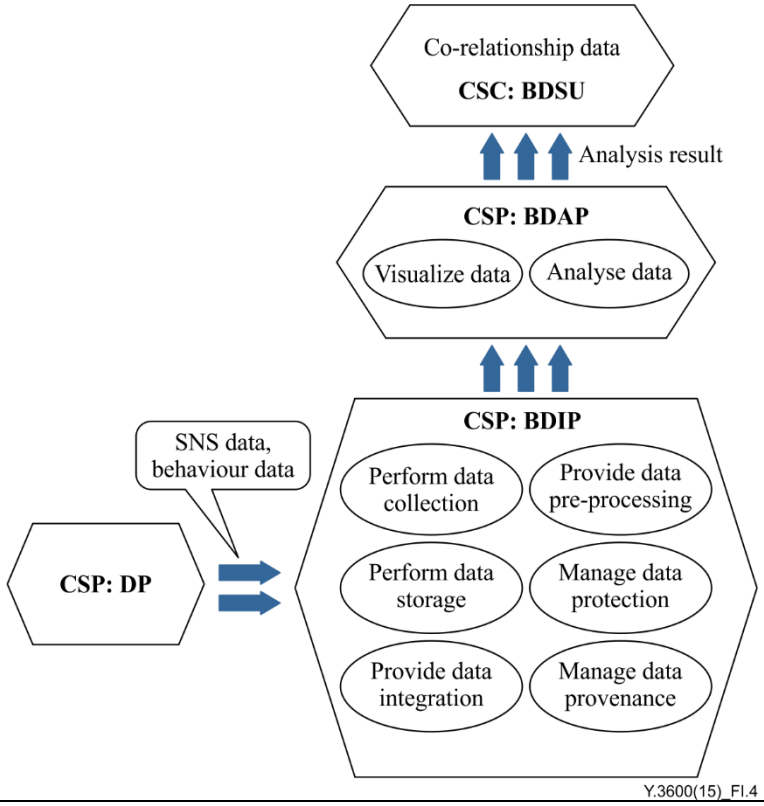
Title	Internet social network services big data analysis
Description	Social network service (SNS) providers use big data application services to analyse the relationships between social network service users. Social network service users register to SNS services with parts of their background information, such as education and working experience, professional skills, etc. The data provider (SNS platform provider) saves the registered users information in databases. The SNS platform provider uses big data services provided by the CSP:BDIP and CSP:BDAP who use cloud computing systems and technologies, such as distributed databases, computing frameworks, cache systems, etc. The SNS platform provider could analyse the behaviour of the SNS users and their co-relationships between each other. The CSC:BDSU (SNS platform provider) could push and promote the contact information to the SNS users that the pushed persons may have some relationship with receivers through pop messages or emails. Privacy rules are followed, if applicable.
Roles/sub-roles	– CSN:DP – CSP:BDAP – CSP:BDIP – CSC:BDSU
Name	
Pre-conditions	– The CSN:DP can provide the raw data reflecting the behaviours of the SNS users.
Post-conditions	– The CSP:BDAP can have responsibility for providing analysis results of network user behaviour to the CSC:BDSU. It can have responsibility for providing the co-relationships between the users of an SNS. The CSP:BDAP should follow the privacy rules, if applicable. – The CSC:BDSU utilizes data of co-relationships between the users of an SNS to give recommendations of new services or new friends in an SNS, etc., following privacy rules, if applicable.
Derived requirements	– Data storage (refer to clause 8.3, requirements 1, 2, 5, clause 8.7, requirement 1) – Data analysis (refer to clause 8.4, requirements 6, 7)

Table I.5 – Mobile network user behaviour big data analysis

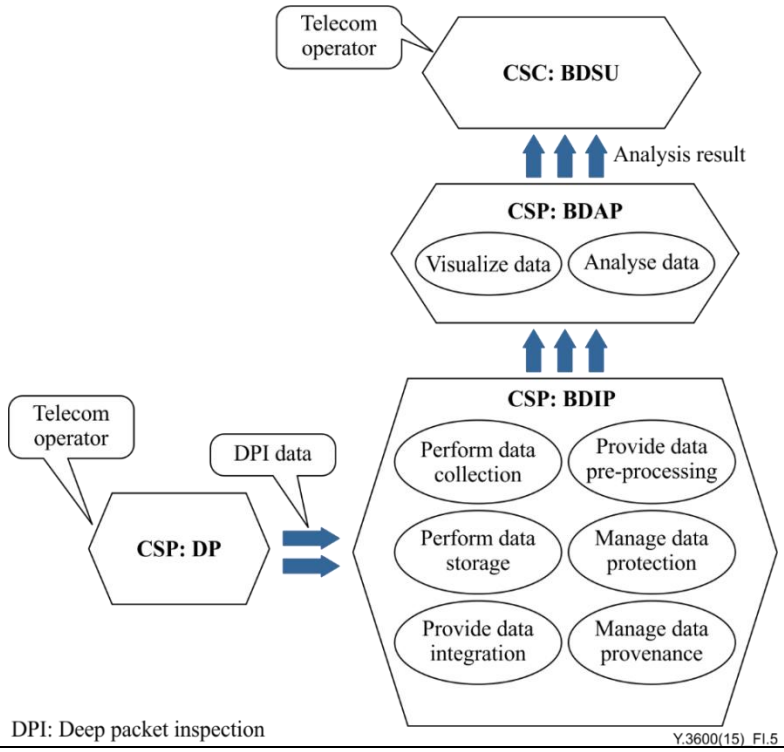
Title	Mobile network user behaviour big data analysis
Description	Telecom operators use big data application services to analyse Internet accessing behaviour of mobile network users. Telecom mobile network users use a 3G/4G network to access mobile Internet services (web browsing, social networking, e-commerce, music, video, mobile television, etc.) using mobile devices, such as mobile phones, tablet computers, PDAs and so on. Optical splitter and deep packet inspection (DPI) devices have been deployed along with the PDSN or serving general packet radio service support node (SGSN) devices. DPI devices capture the data packets of mobile network devices from the optical splitter and extract the Internet access attributes information. Telecom operators set up private big data infrastructure using cloud computing technologies or buy services from the CSP:BDIP and build big data application services or buy services from the CSP:BDAP. Telecom operators could push and promote value added services, applications and advertisements for suitable mobile network users such as a CSC:BDSU. Privacy rules are followed, if applicable.
Roles/sub-roles	– CSN:DP – CSP:BDAP – CSP:BDIP – CSC:BDSU
Figure	
Pre-conditions	– The CSN:DP can provide the raw data which reflects mobile network users' behaviour.
Post-conditions	– The CSP:BDAP can have responsibility for providing analysis results of mobile network user behaviour to the CSC:BDSU. The analysis results show the users' habits of using network services, such as in which time slots the users may access which websites or applications. – The CSC:BDSU pushes and promotes the value added services, applications and advertisements for suitable mobile network users.
Derived requirements	– Data collection (refer to clause 8.1, requirement 1) – Data storage (refer to clause 8.3, requirement 1) – Data analysis (refer to clause 8.4, requirements 2, 7, 12) – Data application (refer to clause 8.4, requirement 13, clause 8.7, requirement 1)

Table I.6 – Big data based mobile network (3G/4G) billing detail query

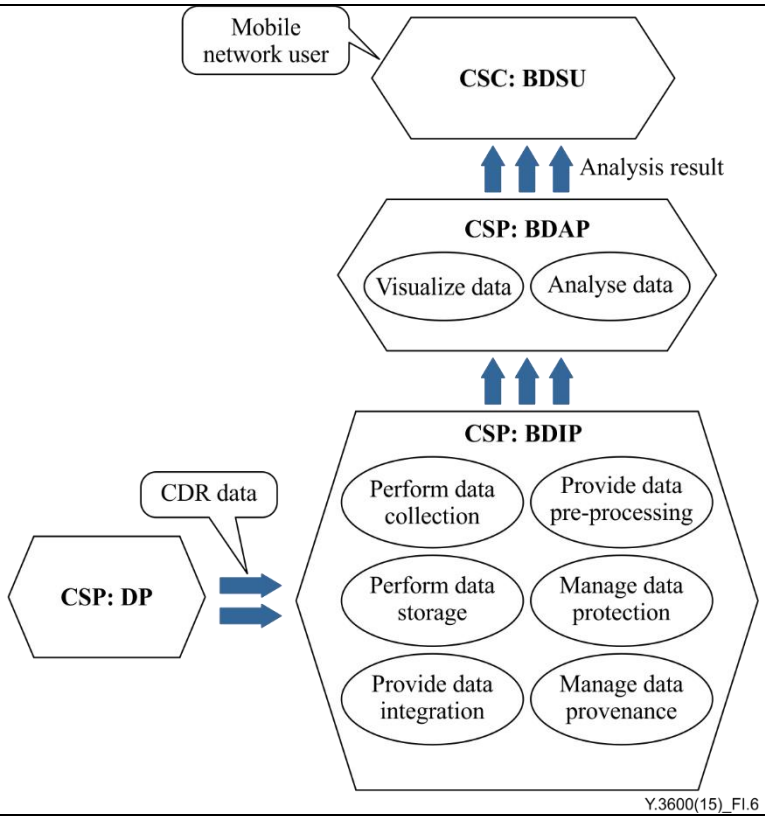
Title	Big data based mobile network (3G/4G) billing detail query
Description	Telecom operators use big data application services to provide detailed billing information of mobile network users. Telecom mobile network users use 3G/4G networks to access mobile internet services using mobile devices such as mobile phones and tablet computers .The access network devices, such as PDSN or combined GPRS service node (SGSN/CGSN) devices, record all kinds of billing information of each of the mobile network users into a standardized format, such as a charging detail record (CDR). A charging gateway function (CGF) modular collects different charging records and forwards these charging records to the billing system. The billing system keeps and processes all the CDRs from different CGFs and finally generates the charging bill for each mobile user. Telecom operators set up big data infrastructure and store huge amounts of user billing detail information. Mobile network users could quickly query the detailed billing information based on the big data infrastructure and big data applications. A mobile network user could get the statistics and analytics information for instance in terms of time, type of website and applications accessed through a 3G/4G mobile network for his/her internet access actions through 3G/4G mobile networks.. Privacy rules are followed, if applicable.
Roles/sub-roles	– CSN:DP – CSP:BDAP – CSP:BDIP – CSC:BDSU
Figure	
Pre-conditions	– CSN:DP can provide CDR data of network users

Table I.6 – Big data based mobile network (3G/4G) billing detail query

Post-conditions	– CSP:BDAP can have functionalities for providing analysis result of network users' CDR data to the CSC:BDSU. – CSC:BDSU could quickly query the detailed billing record for each internet access based on the analysis result.
Derived requirements	– Data collection (refer to clause 8.1, requirement 3) – Data storage (refer to clause 8.3, requirement 4) – Data analysis (refer to clause 8.4, requirement 8) – Data application (refer to clause 8.5, requirement 2)

Table I.7 – Intelligent transport big data analysis

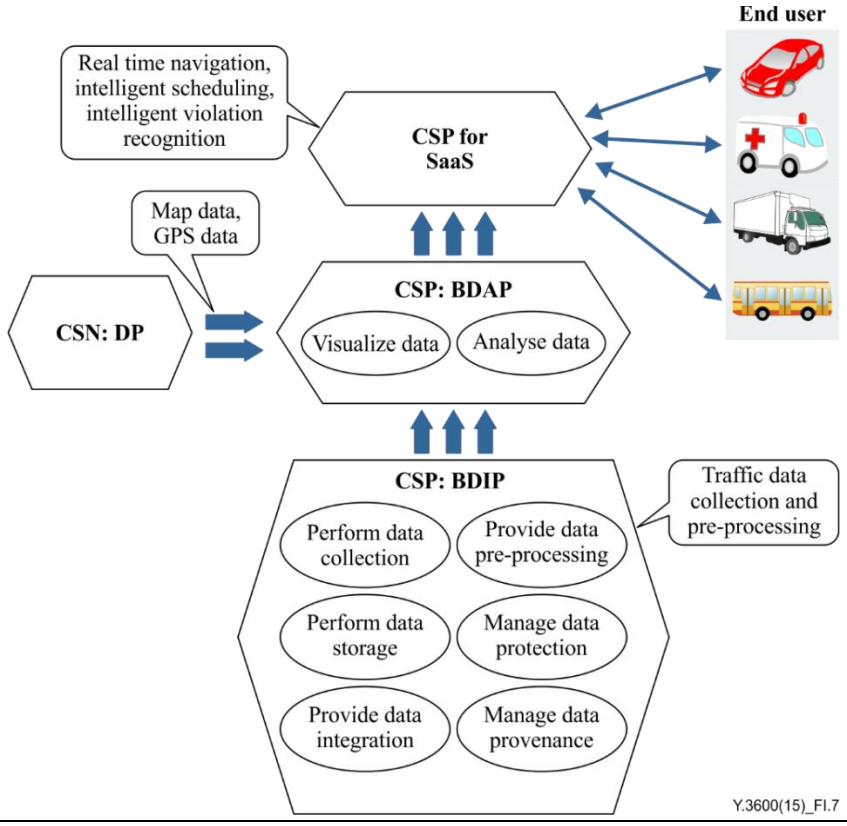
Title	Intelligent transport big data analysis
Description	The public transportation traffic information on the main roads is collected by an intelligent transport platform. Intelligent transport platform providers use big data applications to implement backend supporting services such as traffic volume prediction, route traffic analysis and route navigation optimization.
Roles/sub-roles	– CSN:DP – CSP:BDAP – CSP:BDIP – CSC:BDSU
Figure	
Pre-conditions	– The speed sensors and high-resolution cameras are installed at intersections and key points of roads in the city and outskirts. – The traffic of vehicles and the status of roads data are collected in real time. – The big data applications for intelligent transport systems are built upon the private or public big data infrastructure.

Table I.7 – Intelligent transport big data analysis

Post-conditions	– Real time and accurate route navigation based on the shortest path algorithm and route traffic could be provided to private and business vehicle drivers of all types of vehicles on the road except drivers of track vehicles. – Accurate scheduling for bus departures is established on the real time analysis and prediction of road traffic. – Automatic recognition of violations of road traffic laws such as speeding, running red lights, drink-driving, running through restricted areas, fatigue driving, etc.
Derived requirements	– Data analysis (refer to clause 8.4, requirement 3, 9,12)

Appendix II

Use cases of cloud computing based big data as analysis services

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

Table II.1 – Continuous product improvement

Title	Continuous product improvement
Description	A software game manufacturing company would like to concentrate on application development as a basic activity and the area of investment. In order to continuously improve its products, the company needs an analytics solution to collect and analyse the associated market and users' feedback. The company embedded monitoring functions in the software games to observe users' behaviours, activities and preferences of available game options. In addition a questionnaire is included in the game application for obtaining offline users' feedback. Other sources of information about the game perception are Internet forums, discussion lists and social communities, as well as press and television publications. Given the large number of game users, the company has a lot of information to sift through every month. But the amount of data can change depending on the schedule of new games publication, so proper planning and smart organization of feedback data analysis would help the company's IT team in their use of big data analytics to improve business success. The game company builds an analytical solution based on big data as a service (BDaaS) by defining: – the schedule of BDaaS activity depending on games publication time; – requirements for data collection directly from the application using embedded monitoring functions; – requirements for collection of Internet and press data related to the game in order to analyse users' and market opinions and its evolution; – requirements for social media data analysis of game users' opinions and direct exchange between users; – the access method to the user generated questionnaire data about the game; – database selection for all data collected about the game; – requirements for analytics and reporting to emphasize the opinions about the application, identify strong and weak points, gaps and spare elements; – monitoring criteria to analyse the intensity of new opinions on appearance generated by users to identify the application maturity level, which allows decisions to be made about ending involvement in the product; – performance objectives to be reached by the BDaaS system. The game company orders a BDaaS including the following elements: – specification of data sources, which need to be serviced; – data integration goals for overall analytics; – database functionalities; – data storage functionalities (data types, variety, lifetime); – analytical tool functionalities; – access to analytical tool configuration and customization; – reporting schemes and integration with the company internal systems; – data collected and results of analysis available only for the company, even if publicly available data are used.

Table II.1 – Continuous product improvement

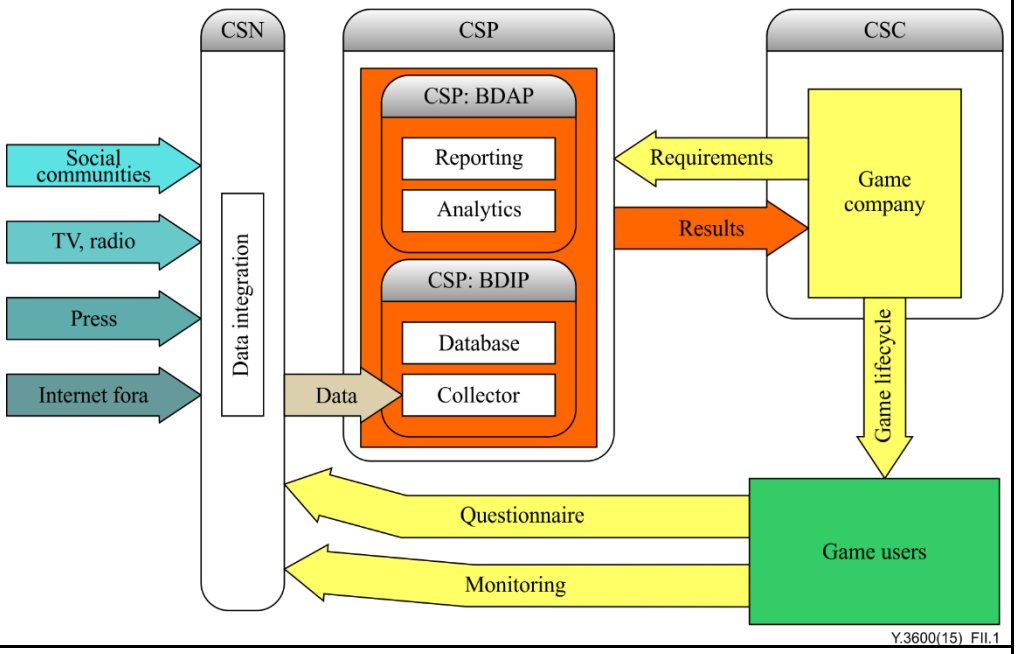
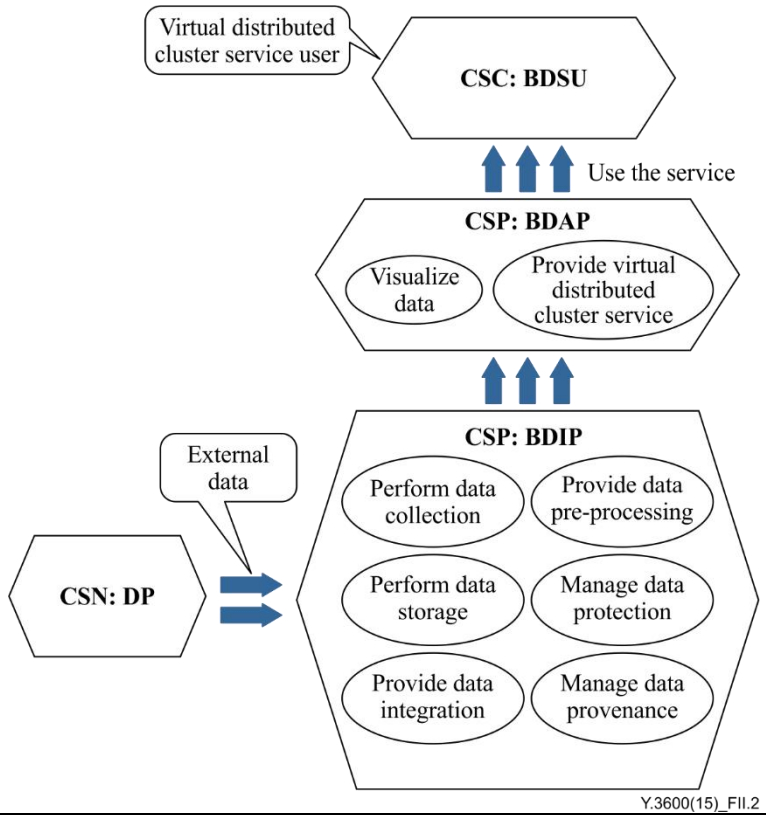
Roles/sub-roles	– The cloud service provider should support the CSP:BDIP role to collect, process, store and delete data according to the CSC (i.e., the game manufacturing company) specifications. In addition, the CSP should support the CSP:BDAP role to provide analytical applications with integration of CSC systems, as well as reporting solutions.
Figure	
Pre-conditions (optional)	The game company has to plan the analytical solution to support application development. The solution utilization can vary depending on the new products being launched, so the investment in a solution may prove to be ineffective.
Post-conditions (optional)	The game company can define analytical campaigns for each product and optimize the analysis effectiveness. Only currently required capabilities are used from the BDaaS solution in a pay-as-you-go model.
Derived requirements	– Data collection (clause 8.1, requirements 1, 3, 5) – Data pre-processing (clause 8.2, requirements 3, 4) – Data storage (clause 8.3, requirements 1, 2, 3, 4, 6, 7) – Data analysis (clause 8.4, requirements 5, 10, 11) – Data visualization (clause 8.5, requirements 1, 3, 4, 5) – Data security and protection (clause 8.7, requirement 4)

Table II.2 – Virtualized distributed cluster service

Title	Virtualized distributed cluster service
Description	A virtual distributed cluster service is a typical web service which makes it easy to have a cluster of machines quickly and cost-effectively processing vast amounts of data provided by a CSP. The virtual distributed cluster service uses distributed clustering software as a framework, to distribute the customers' data and processing across a resizable cluster of virtual machine instances in cloud resource pools. Three steps that are often included by the CSC using a virtual cluster service are: 1) **Upload data**. The CSC:BDSU uploads the data that needs to be analysed to the cloud storage space that belongs to the CSC:BDSU. In addition the CSC:BDSU could use the data provided by the CSN:DP.

Table II.2 – Virtualized distributed cluster service

	2) **Create virtual distributed cluster**. The CSC:BDSU creates and configures the distributed cluster by specifying data inputs, outputs, cluster size, security settings and other necessary parameters. 3) **Monitor and collect**. CSC:BDSU monitors the health and progress of the distributed cluster using the tools provided by the CSP. When the distributed processing job is completed the CSC:BDSU retrieves the output in the specified storage space. A virtual distributed cluster service could be used in a variety of applications, including log analysis, web indexing, data warehousing, machine learning, financial analysis, scientific simulation and bioinformatics.
Roles/Sub-roles	– CSN:DP – CSP:BDAP – CSP:BDIP – CSC:BDSU
Figure	
Pre-conditions (optional)	– The CSC:BDSU has registered in the service platform of the CSP of the virtual distributed cluster service. – The CSC:BDSU has applied the cloud storage space for accepting the input data and output data from the virtual distributed cluster service cluster.
Post-conditions (optional)	– The CSC:BDSU could quickly create the specified cluster with the desired number of virtual machines and software versions. – The CSC:BDSU gets the processing results after a planned time period.
Derived requirement	– Data storage (refer to clause 8.3, requirement 1) – Data analysis (refer to clause 8.4, requirements 2, 4, 7) – Data management (refer to clause 8.6, requirements 1, 2, 3, 4, 5, 6) – Data security and protection (refer to clause 8.7, requirements 2, 3, 4)

Appendix III

Mapping of big data ecosystem roles into user view of ITU-T Y.3502

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

Table III.1 shows the results of mapping between the sub-role of the cloud computing reference architecture (CCRA) user view and the performance of similar roles in the big data ecosystem. A similar sub-role with a data supplier does not exist in CCRA. A big data service customer and a CSC may be considered to perform similar activities. Similarly, a data broker and CSN: cloud service broker may be considered in the same manner. If the CCRA is assumed to contain the big data service area, the activities for the CSN: cloud service broker should be extended in the data perspective. The big data service provider is related to all sub-roles of the CSP. This means that the big data service provider could be treated with an implanted feature or an independent sub-role of the CSP.

Table III.1 – Mapping of big data ecosystem roles and sub-roles of the CCRA user view

Big data ecosystem		**User view of ITU-T Y.3502**	**Note**
Data provider	Data supplier	CSN	For the cloud computing based big data environment, the new sub-role of the CSN for the data provider is required.
	Data broker	CSN	For the cloud computing based big data environment: (option 1) the extension of the activities of the CSN: Cloud service broker is required. (option 2) the new sub-role of the CSN for data brokerage is required. (option 3) adding the "brokerage" activity on the sub-role which corresponds with the data supplier is required.
Big data service customer		CSC	For the cloud-based big data service environment: (option 1) using the CSC is required. (option 2) the new role or sub-role of the CSC for the big data service customer is required.
Big data service provider		CSP	The big data service provider could be treated with an implemented feature. Nevertheless, to clarify the cloud-based big data service, an independent sub-role of the CSP for the big data service provider is required.

Bibliography

[b-ITU-T M.3030] Recommendation ITU-T M.3030 (2002), *Telecommunications Markup Language (tML) framework.*

[b-ITU-T Y.2201] Recommendation ITU-T Y.2201 (2011), *Requirements and capabilities for ITU-T NGN*.

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

[b-CRA-BDWP] *Challenges and Opportunities with Big Data*, Computing Research Association, November 2012.
<http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/bigdatawhitepaper.pdf>

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 and next-generation networks**
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2015