

Recommendation

ITU-T Y.3607 (01/2023)

SERIES Y: Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities

Big Data

Big data – Functional architecture for data provenance

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, 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
Computing power networks	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.3599
-----------------	---------------

BIG DATA	Y.3600–Y.3799
-----------------	----------------------

QUANTUM KEY DISTRIBUTION NETWORKS	Y.3800–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
Evaluation and assessment	Y.4900–Y.4999

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

Recommendation ITU-T Y.3607

Big data – Functional architecture for data provenance

Summary

Recommendation ITU-T Y.3607 describes a functional architecture for big data provenance (BDP). To provide the functional architecture for big data provenance, the big data provenance functions are defined based on the functional requirements and logical components identified in Recommendation ITU-T Y.3602. This Recommendation also provides the relationship between the functional architecture of big data provenance and the big data reference architecture in Recommendation ITU-T Y.3605.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y.3607	2023-01-13	13	11.1002/1000/15242

Keywords

Big data, data provenance, functional architecture.

* 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/software copyrights, 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 appropriate ITU-T databases available via the ITU-T website at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2023

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope.....	1
2	References.....	1
3	Definitions	1
	3.1 Terms defined elsewhere	1
	3.2 Terms defined in this Recommendation.....	2
4	Abbreviations and acronyms	2
5	Conventions	3
6	Relationship between BDP logical components and BDP functions	3
7	BDP functions.....	3
	7.1 Provenance unit processing	3
	7.2 Provenance information processing	4
	7.3 Big data analytics support	5
	7.4 Big data system connection	7
8	BDP functional architecture	7
	8.1 Provenance unit processing functions	8
	8.2 Provenance information processing functions.....	9
	8.3 Big data analytics support functions.....	10
	8.4 Big data system connection functions	11
9	Reference points	11
	9.1 Reference points between PUP-FS and BDSC-FS.....	11
	9.2 Reference point between PUP-FS and PIP-FS	11
	9.3 Reference point between PUP-FS and BDAS-FS	12
	9.4 Reference point between PIP-FS and BDSC-FS.....	12
	9.5 Reference points between BDAS-FS and BDSC-FS	12
	9.6 Reference points within PUP-FS	12
	9.7 Reference points within PIP-FS	12
	9.8 Reference points within BDAS-FS	12
10	Security considerations	13
	Appendix I – Relationships among BDP functional requirements, BDP logical components, and BDP functions.....	14
	Appendix II – The overall relationships among BDP functions.....	20
	Appendix III – Relationships between the BDP functions and functional components of big data architecture in [ITU-T Y.3605].....	21
	Bibliography.....	22

Recommendation ITU-T Y.3607

Big data – Functional architecture for data provenance

1 Scope

This Recommendation provides a functional architecture for big data provenance. It specifies the following:

- Functions for supporting big data provenance;
- Functional architecture of big data provenance;
- Reference points among functions for big data provenance.

In addition, this Recommendation provides a relationship between the functional architecture of big data provenance (BDP) and the big data reference architecture of [ITU-T Y.3605] in appendices.

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 Y.3602] Recommendation ITU-T Y.3602 (2022), *Big data – Functional requirements for data provenance*.

[ITU-T Y.3605] Recommendation ITU-T Y.3605 (2020), *Big data – Reference architecture*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 big data [b-ITU-T Y.3600]: 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.

3.1.2 big data provenance [ITU-T Y.3602]: Information that records the historical path of data according to the data lifecycle operations in a big data ecosystem.

NOTE 1 – Data lifecycle operations include data generation, transmission, storage, use and deletion.

NOTE 2 – Data provenance information provides details about the source of data, such as the person responsible for the provision of data, functions applied to data, and information about the computing environment for data processing (e.g., operating system, description of the hardware, locale settings and time zone).

3.1.3 provenance [b-ITU-T X.1255]: Information pertaining to any source of information including the party or parties involved in generating it, introducing it, and/or vouching for it.

3.1.4 reference architecture [b-ISO/IEC 26550]: Core architecture that captures the high-level design of a software and systems product line including the architectural structure and texture

(e.g., common rules and constraints) that constrains all member products within a software and systems product line.

NOTE – Application architectures of the member products included in the product line reuses (possibly with modifications) the common parts and binds variable parts of the domain architecture. Application architectures of the member products may (but do not need to) provide variability.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

BD	Big Data
BDAS-FS	Big Data Analytics Support Functions
BDP	Big Data Provenance
BDSC-FS	Big Data System Connection Functions
BDSI-F	Big Data System Interface Function
BDSP	Big Data Service Provider
DB	Data Broker
DP	Data Provider
H/W	Hardware
JSON	JavaScript Object Notation
OS	Operating System
PIC-F	Provenance Information Composition Function
PII	Personally Identifiable Information
PIM-F	Provenance Information Monitor Function
PIP-FS	Provenance Information Processing Functions
PIPM-F	Provenance Information Policy Management Function
PIT-F	Provenance Information Transformation Function
PUE-F	Provenance Unit Extraction Function
PUM-F	Provenance Unit Management Function
PUP-FS	Provenance Unit Processing Function
PUV-F	Provenance Unit Validator Function
UIAS-F	User Interface for Analytic Support Function
WE-F	Workflow Explorer Function
WMB-F	Workflow Map Building Function
WM-F	Workflow Management Function
XML	extensible Markup Language

5 Conventions

Throughout this Recommendation, the term "big data system" is to be understood as a big data system that includes capabilities of big data provenance functions.

6 Relationship between BDP logical components and BDP functions

There are six big data provenance (BDP) logical components in [ITU-T Y.3602] including:

- 1) provenance model management;
- 2) provenance lifecycle management;
- 3) provenance sharing policy management;
- 4) personally identifiable information (PII) management;
- 5) analysis support; and
- 6) monitoring.

These BDP logical components are realized by the BDP functions. Table 6-1 shows the relationship between BDP logical components in [ITU-T Y.3602] and BDP functions in this Recommendation.

Table 6-1 – Relationship between BDP logical components and BDP functions

BDP logical components in [ITU-T Y.3602]	BDP functions in this Recommendation
Provenance model management	Provenance unit processing (see clause 7.1)
Provenance lifecycle management	Provenance unit processing (see clause 7.1)
Provenance sharing policy management	Provenance information processing (see clause 7.2)
Personally identifiable information (PII) management	Provenance information processing (see clause 7.2)
Analysis support	Big data analytics support (see clause 7.3), and big data system connection (see clause 7.4)
Monitoring	Provenance information processing (see clause 7.2), and big data system connection (see clause 7.4)

7 BDP functions

This clause provides the BDP functions which are derived from the BDP functional requirements in [ITU-T Y.3602]. The relationships between BDP functional requirements and BDP functions are described in Appendix I.

7.1 Provenance unit processing

The provenance unit processing functions (PUP-FS) include:

- provenance unit extraction (see clause 7.1.1);
- provenance unit validator (see clause 7.1.2);
- provenance unit management (see clause 7.1.3).

7.1.1 Provenance unit extraction

The provenance unit extraction function (PUE-F) extracts the provenance units from the big data (BD) system. This function: