INTERNATIONAL STANDARD

ISO/IEC 20547-3

First edition
2020-03

Information technology — Big data reference architecture —

Part 3: Reference architecture

Technologies de l'information — Architecture de référence des mégadonnées —

Partie 3: Architecture de référence

Reference number
ISO/IEC 20547-3:2020(E)

© ISO/IEC 2020



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2020

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **2**
4 Abbreviated terms **4**
5 Conventions **5**
6 Big data reference architecture concepts **5**
6.1 General 5
6.2 Views 6
6.3 Overview of user view 6
6.4 Overview of functional view 7
6.5 Relationship between the user view and the functional view 8
6.6 Relationship of the user view and functional view to cross-cutting aspects 8
7 User view **9**
7.1 Big data roles, sub-roles, and activities 9
7.2 Role: Big data application provider (BDAP) 10
7.2.1 General 10
7.2.2 Sub-role: big data collection provider (BDCP) 11
7.2.3 Sub-role: big data preparation provider (BDPreP) 11
7.2.4 Sub-role: big data analytics provider (BDAnP) 12
7.2.5 Sub-role: big data visualization provider (BDVP) 12
7.2.6 Sub-role: big data access provider (BDAcP) 12
7.3 Role: big data framework provider (BDFP) 12
7.3.1 General 12
7.3.2 Sub-role: big data infrastructure provider (BDIP) 13
7.3.3 Sub-role: big data platform provider (BDPlaP) 13
7.3.4 Sub-role: big data processing provider (BDProP) 13
7.4 Role: big data service partner (BDSP) 14
7.4.1 General 14
7.4.2 Sub-role: big data service developer (BDSD) 15
7.4.3 Sub-role: big data auditor (BDA) 15
7.4.4 Sub-role: big data system orchestrator (BDSO) 15
7.5 Role: big data provider (BDP) 16
7.6 Role: big data consumer (BDC) 16
8 Cross-cutting aspects **17**
8.1 General 17
8.2 Security and privacy 17
8.3 Management 17
8.4 Data governance 18
9 Functional view **18**
9.1 Functional architecture 18
9.1.1 General 18
9.1.2 Layering architecture 19
9.1.3 Multi-layer functions 20
9.2 Functional components 20
9.2.1 General 20
9.2.2 Big data application layer functional components 21
9.2.3 Big data processing layer functional components 23
9.2.4 Big data platform layer functional components 25
9.2.5 Resource layer functional components 28

© ISO/IEC 2020 – All rights reserved

9.2.6 Multi-layer functional components 29
Annex A (informative) **Mapping big data RA functional view to other system integration RA** **33**
Annex B (informative) **Examples of the relationship of roles in big data ecosystem** **34**
Annex C (informative) **35**
Bibliography **37**

 © ISO/IEC 2020 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

A list of all parts in the ISO/IEC 20547 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

© ISO/IEC 2020 – All rights reserved

Introduction

The ISO/IEC 20547 series is intended to provide users with a standardized approach to developing and implementing big data architectures and provide references for approaches. ISO/IEC TR 20547-1 provides users with an overview of the reference architecture framework described in this document and a process for applying that framework in developing an architecture. ISO/IEC TR 20547-2 provides a collection of big data use cases and decomposes those use cases into technical considerations that big data architects and system implementers can consider. This document describes the reference architecture in terms of User and Functional views. Those views can be used by the big data architect to describe their specific system. ISO/IEC 20547-4 describes the security and privacy aspects unique to big data. ISO/IEC TR 20547-5 provides a list of standards and their relationship to the reference architecture that architects and implementers can consider as part of the design and implementation of their system.

Each of these parts is built on the common vocabulary and concepts described in ISO/IEC 20546.

In general terms, reference architecture provides an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions (see 3.2). Reference architectures generally serve as a reference foundation for solution architectures and can also be used for comparison and alignment purposes.

The key goal of this reference architecture is to facilitate a shared understanding across multiple products, organizations, and disciplines about current architectures and future direction.

The reference architecture presented in this document provides an architecture framework for describing the big data components, processes, and systems to establish a common language for the various stakeholders named as big data reference architecture (BDRA). It does not represent the system architecture of a specific big data system. Instead, it is a tool for describing, discussing, and developing system-specific architectures using an architecture framework of reference. It provides generic high-level architectural views that are an effective tool for discussing the requirements, structures, and operations inherent to big data. The model is not tied to any specific vendor products, services or reference implementation, nor does it define prescriptive solutions that inhibit innovation.

© ISO/IEC 2020 – All rights reserved

Information technology — Big data reference architecture —

Part 3: Reference architecture

1 Scope

This document specifies the big data reference architecture (BDRA). The reference architecture includes concepts and architectural views.

The reference architecture specified in this document defines two architectural viewpoints:

— a user view defining roles/sub-roles, their relationships, and types of activities within a big data ecosystem;

— a functional view defining the architectural layers and the classes of functional components within those layers that implement the activities of the roles/sub-roles within the user view.

The BDRA is intended to:

— provide a common language for the various stakeholders;

— encourage adherence to common standards, specifications, and patterns;

— provide consistency of implementation of technology to solve similar problem sets;

— facilitate the understanding of the operational intricacies in big data;

— illustrate and understand the various big data components, processes, and systems, in the context of an overall big data conceptual model;

— provide a technical reference for government departments, agencies and other consumers to understand, discuss, categorize and compare big data solutions; and

— facilitate the analysis of candidate standards for interoperability, portability, reusability, and extendibility.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8000-2, *Data quality — Part 2: Vocabulary*

ISO/TS 8000-60, *Data quality — Part 60: Data quality management: Overview*

ISO 8000-61, *Data quality — Part 61: Data quality management: Process reference model*

ISO/IEC 38500, *Information technology — Governance of IT for the organization*

ISO/IEC 38505-1, *Information technology — Governance of IT — Governance of data — Part 1: Application of ISO/IEC 38500 to the governance of data*

© ISO/IEC 2020 – All rights reserved

ISO/IEC TR 38505-2, *Information technology — Governance of IT — Governance of data — Part 2: Implications of ISO/IEC 38505-1 for data management*

ISO 55000, *Asset management — Overview, principles and terminology*

ISO 55001, *Asset management — Management systems — Requirements*

ISO 55002, *Asset management — Management systems — Guidelines for the application of ISO 55001*

ISO/IEC/IEEE 42010, *Systems and software engineering — Architecture description*

ISO/IEC 20546, *Information technology — Big data — Overview and vocabulary*

ISO/IEC 17789, *Information technology — Cloud computing — Reference architecture*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8000-2, ISO/TS 8000-60, ISO 8000-61, ISO/IEC 38500, ISO/IEC 38505-1, ISO/IEC TR 38505-2, ISO 55000, ISO 55001, ISO 55002, ISO/IEC/IEEE 42010, ISO/IEC 20546, ISO/IEC 17789 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
data
reinterpretable representation of *information* (3.3) in a formalized manner suitable for communication, interpretation, or processing

[SOURCE: ISO/IEC 2382:2015, 2121272]

3.2
reference architecture
authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions

Note 1 to entry: This document utilizes the definition of reference architecture from DoD "reference architecture description"[Z].

Note 2 to entry: Reference architectures generally serve as a foundation for solution architectures and can also be used for comparison and alignment of instantiations of architectures and solutions.

3.3
information
data (3.1) that are processed, organized and correlated to produce meaning

Note 1 to entry: Information concerns facts, concepts, objects, events, ideas, processes, etc.

3.4
activity
specified pursuit or set of tasks

[SOURCE: ISO/IEC 17789:2014, 3.2.1]

3.5
knowledge
maintained, processed, and interpreted *information* (3.3)

[SOURCE: ISO 5127:2017, 3.1.1.17]

© ISO/IEC 2020 – All rights reserved

3.6
functional component
functional building block needed to engage in an *activity* (3.4), backed by an implementation

[SOURCE: ISO/IEC 17789:2014, 3.2.3]

3.7
data governance
property or ability that needs to be coordinated and implemented by a set of *activities* (3.4) aimed to design, implement and monitoring a *strategic plan for data asset management*

Note 1 to entry: Governance of data is described in ISO/IEC 38505-1.

Note 2 to entry: Data asset is understood as a set of data items, or data entities, that have a real or potential benefit for an organization. Data asset is a subset of asset defined in ISO 55000. A benefit is an advantage to the organization of the actionable knowledge derived from an analytic system. It is often ascribed to big data due to the understanding that data has potential benefit that was typically not considered previously.

Note 3 to entry: A strategic plan for data asset management is a document specifying how *data management* (3.15) is to be aligned to the organizational strategy. This term has the same meaning as strategic asset management plan (SAMP) defined in ISO 55000 with data point of view.

3.8
data quality
degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions

[SOURCE: ISO/IEC 25024:2015, 4.11]

3.9
data quality management
coordinated activities to direct and control an organization with regard to data quality

[SOURCE: ISO 8000-2:2018, 3.4.9]

3.10
party
natural person or legal person, whether or not incorporated, or a group of either

[SOURCE: ISO/IEC 17789:2014, 7.2.3]

3.11
policy
intention and direction of an organization as formally expressed by its top management

[SOURCE: ISO 55000:2014, 3.1.18, modified — The term has been changed to the singular form and the final stop has been removed from the definition.]

3.12
role
set of *activities* (3.4) that serves a common purpose

[SOURCE: ISO/IEC 17789:2014, 3.2.7]

3.13
stream
list of flow objects attached to a port of a flow object

[SOURCE: ISO/IEC 10179:1996, 4.33, modified — by deleting leading article and trailing full stop.]

© ISO/IEC 2020 – All rights reserved

3.14
sub-role
subset of the *activities* (3.4) of a given *role* (3.12)

[SOURCE: ISO/IEC 17789:2014, 3.2.9]

3.15
data management
set of *activities* (3.4) aimed to implement the big data architecture that best meet business goals by following the strategic plan for data management assessment

3.16
data lifecycle
stages in the management of a data

Note 1 to entry: The target of lifecycle (defined in ISO 55000) is data in this document.

3.17
application programming interface
API
boundary across which application software uses facilities of programming languages to invoke services

[SOURCE: ISO/IEC 18012-2:2012, 3.1.4, modified — Note 1 to entry has been removed and the final stop has been deleted from the definition.]

4 Abbreviated terms

ACID	atomicity, consistency, isolation, and durability
API	application programming interface
CEP	complex event processing
CPU	central processing unit
BDA	big data auditor
BDAP	big data application provider
BDAcP	big data access provider
BDAnP	big data analytics provider
BDC	big data consumer
BDCP	big data collection provider
BDFP	big data framework provider
BDIP	big data infrastructure provider
BDP	big data provider
BDPlaP	big data platform provider
BDPreP	big data preparation provider
BDProP	big data processing provider

© ISO/IEC 2020 – All rights reserved