TECHNICAL REPORT

ISO/IEC TR 20547-1

First edition
2020-08

Information technology — Big data reference architecture —

Part 1: Framework and application process

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

Partie 1: Cadre méthodologique et processus d'application

Reference number
ISO/IEC TR 20547-1:2020(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
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ..... **iv**
Introduction ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Abbreviated terms **2**
5 Document overview **3**
6 Big data standardization: motivation and objectives **3**
7 Conceptual foundations **5**
7.1 General 5
7.2 Reference architecture concepts 5
7.3 Reference architecture structure 6
8 Big data reference architecture elements **7**
8.1 Overview 7
8.2 Stakeholders 8
8.3 Concerns 9
8.4 Views 9
8.4.1 User view 10
8.4.2 Functional view 10
9 Big data reference architecture application process **10**
9.1 Overview 10
9.2 Identify stakeholders and concerns 11
9.3 Map stakeholders and concerns to roles and subroles 11
9.4 Develop detailed activity descriptions and map to concerns 12
9.5 Define functional components to implement activities 13
9.6 Cross walk activities/functional components back to concerns 13
Bibliography **14**

© 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.

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) or the IEC list of patent declarations received (see http://patents.iec.ch).

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 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 big data paradigm is a rapidly changing field with rapidly changing technologies. This dynamic situation creates two significant issues for potential implementers of the technology. First, there is a lack of standard definitions for terms including the core concept of big data. The second issue is that there is no consistent approach to describe a big data architecture and implementation. The first issue is addressed by ISO/IEC 20546. The ISO/IEC 20547 series is targeted to the second issue and provides a framework and reference architecture which organizations can apply to their problem domain to effectively and consistently describe their architecture and its implementations with respect to the roles/actors and their concerns as well as the underlying technology. This document describes the reference architecture framework and provides a process for mapping a specific problem set/use case to the architecture and evaluating that mapping.

© ISO/IEC 2020 – All rights reserved

Information technology — Big data reference architecture —

Part 1: Framework and application process

1 Scope

This document describes the framework of the big data reference architecture and the process for how a user of the document can apply it to their particular problem domain.

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

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC/IEEE 42010 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
big data
extensive datasets — primarily in the characteristics of volume, variety, velocity, and/or variability — that require a scalable technology for efficient storage, manipulation, and analysis

Note 1 to entry: Big data is commonly used in many different ways, for example as the name of the scalable technology used to handle big data extensive datasets.

[SOURCE: ISO/IEC 20546:2019, 3.1.2]

3.2
reference architecture
in the field of software architecture or enterprise architecture, provides a proven template solution for an architecture for a particular domain, as well as a common vocabulary with which to discuss implementations, often with the aim of stressing commonality

[SOURCE: ISO/TR 14639-2:2014, 2.65]

3.3
framework
particular set of beliefs, or ideas referred to in order to describe a scenario or solve a problem

[SOURCE: ISO 15638-6:2014, 4.30]

© ISO/IEC 2020 – All rights reserved

3.4
security
protection against intentional subversion or forced failure. A composite of four attributes — confidentiality, integrity, availability, and accountability — plus aspects of a fifth, usability, all of which have the related issue of their assurance

[SOURCE: ISO/IEC/IEEE 15288:2015, 4.1, 31]

3.5
privacy
right of individuals to control or influence what information related to them may be collected and stored and by whom that information may be disclosed

[SOURCE: ISO/IEC TR 26927:2011, 3.34]

3.6
provenance
information on the place and time of origin, derivation or generation of a resource or a record or proof of authenticity or of past ownership

[SOURCE: ISO/IEC 11179-7:2019, 3.1.10]

3.7
SQL
database language specified by ISO/IEC 9075

Note 1 to entry: SQL is sometimes interpreted to stand for Structured Query Language but that name is not used in the ISO/IEC 9075 series.

[SOURCE: ISO/IEC 20546:2019, 3.1.36]

3.8
lifecycle
evolution of a system, product, service, project or other human-made entity from conception through retirement

[SOURCE: ISO/IEC/IEEE 15288:2015, 4.1.23]

4 Abbreviated terms

BDA	big data auditor
BDAcP	big data access provider
BDAnP	big data analytics provider
BDAP	big data application provider
BDCP	big data collection provider
BDFP	big data framework provider
BDIP	big data infrastructure provider
BDPlaP	big data platform provider
BDPreP	big data preparation provider
BDProP	big data processing provider

© ISO/IEC 2020 – All rights reserved