INTERNATIONAL STANDARD

ISO/IEC 20546

First edition
2019-02

Information technology — Big data — Overview and vocabulary

Technologies de l'information — Mégadonnées — Vue d'ensemble et vocabulaire

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024

Reference number
ISO/IEC 20546:2019(E)

Documents
Document Sharing Hub
© ISO/IEC 2019

Full document details: ISO_IEC 20546 WG_2019

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2019

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

© ISO/IEC 2019 – All rights reserved

Documents
Document Sharing Hub
Full document details: ISO_IEC 20546 WG_2019

Contents

Page

Foreword ..... **iv**
Introduction ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms, definitions and abbreviated terms **1**
3.1 Terms and definitions 1
3.2 Abbreviated terms 6
4 Key characteristics of big data **6**
4.1 General 6
4.2 Key data characteristics 6
4.2.1 Data volume 6
4.2.2 Data velocity 6
4.2.3 Data variety 6
4.2.4 Data variability 6
4.3 Key data processing characteristics 7
4.3.1 Data science 7
4.3.2 Data volatility 7
4.3.3 Data veracity 7
4.3.4 Benefit 7
4.3.5 Data visualization 7
4.3.6 Structured and unstructured data 7
4.3.7 Scaling 7
4.3.8 Distributed file system 8
4.3.9 Distributed data processing 8
4.3.10 Non-relational databases 8
Annex A (informative) **Cross-cutting concepts of big data** **9**
Bibliography **12**

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024

© ISO/IEC 2019 – All rights reserved

Documents
Document Sharing Hub
Full document details: ISO_IEC 20546 WG_2019

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 Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

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.htm

Copyrighted document, no reproduction or circulation
For review by FO on AI in healthcare
Oct 2024

© ISO/IEC 2019 – All rights reserved
Documents
Document Sharing Hub
Full document details: ISO_IEC 20546 WG_2019

Introduction

The big data paradigm is a rapidly changing field with rapidly changing technologies.

The term big data implies datasets that are extensive in volume, velocity, variety and/or variability. The term does not, however, represent data that is simply larger than before, since this has happened on a regular basis for decades. The specific occurrence that has led to the widespread usage of the term big data is that in the mid-2000s, extensive datasets could no longer be handled using extant data systems. The big data techniques represented a shift at that time to use distributed data management and processing through horizontal scaling to achieve the needed performance efficiency at an affordable cost.

In the evolution of data processing systems, there have been a number of times when the need for efficient, cost-effective data analysis has forced a change in existing technologies. For example, the move to a relational model occurred when methods to reliably handle changes to structured data led in the 1980s to the shift to relational databases that modelled relational algebra. That was a fundamental shift in data handling. The revolution in technologies referred to as big data has arisen because the relational model could no longer efficiently handle all the needs for analysis of large and often unstructured datasets. It is not just that data is larger than before, as data has been steadily getting larger for decades. The big data revolution is instead a one-time fundamental shift in architecture towards parallelization, just as the shift to the relational model was a one-time shift. As relational databases evolved to greater efficiencies over decades, so too will big data technologies continue to evolve. Many of the conceptual underpinnings of big data have been around for years, but the years since the mid-2000s have seen an explosion in scaling technologies and their maturation and application to scaled data systems.

The term big data is overloaded in common usage and is used to represent a number of related concepts, in part because several distinct system dimensions are consistently interacting with each other. To understand this revolution, the interplay of the following aspects needs to be considered: the data and processing characteristics of the datasets, the analysis of the datasets, the performance of the systems that handle the data, the business considerations of cost effectiveness, and the new engineering and analysis techniques for distributed data processing using horizontal scaling.

Annex A provides an overview of several concepts from the broader computing domain which are cross-cutting with respect to big data.

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024

© ISO/IEC 2019 – All rights reserved

Documents
Document Sharing Hub
Full document details: ISO_IEC 20546 WG_2019

Copyrighted document, no reproduction or circulation
For review by FG on AI in healthcare
Oct 2024

Documents
Document Sharing Hub

Full document details: ISO_IEC 20546 WG_2019