



中华人民共和国国家标准

GB/T 35295—2017

信息技术 大数据 术语

Information technology—Big data—Terminology

2017-12-29 发布

2018-07-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

目 次

前言	I
1 范围	1
2 术语和定义	1
2.1 大数据及其应用领域术语	1
2.2 密切相关的通用术语	8
参考文献	10
索引	11

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位:中国电子技术标准化研究院、浪潮软件集团有限公司、浪潮(北京)电子信息产业有限公司、国家信息中心、华为技术有限公司、北京数码大方科技股份有限公司、山东省标准化研究院、北京世纪互联宽带数据中心有限公司、清华大学、中科恒源信息科技有限公司、山西天地科技有限公司、大唐软件技术股份有限公司、成都勤智数码科技股份有限公司、微软(中国)有限公司、福建亿榕信息技术有限公司、中国科学院计算机网络信息中心、北京软件和信息服务交易所有限公司、北京华宇软件股份有限公司、四川长虹佳华数字技术有限公司、中兴通讯股份有限公司、四川久远银海软件股份有限公司、北京东方通科技股份有限公司、北明软件股份有限公司、天津南大通用数据技术股份有限公司、重庆云威科技有限公司、广州赛宝认证中心服务有限公司、北京大学。

本标准主要起草人:梅宏、高林、卫凤林、赵菁华、张群、王建华、黄先芝、吴楠、苏志远、宦茂盛、吕欣、杨莉莉、符海芳、吴东亚、齐建军、万洋、隗玉凯、王文飞、周行健、王建民、王朝坤、张星华、莫宏波、苏晓婷、芦效东、周训波、程之明、刘宇峰、吴志刚、许怡娴、苏江文、石松、胡良霖、闵京华、于铁强、胡才勇、高冀鹏、张楠、李剑飞、邹代泉、黄峥、王东、程曦、乔文涛、赵大航、俞立平、汪疆平、王进宏、冀辉、王洪越、李力、刘小茵、李尧、李宁、张展新、王静、李冰、陈海、徐洋、李易昂、赵俊峰。

信息技术 大数据 术语

1 范围

本标准界定了信息技术大数据领域中的常用术语和定义。

本标准适用于大数据领域的科研、教学和应用。

2 术语和定义

2.1 大数据及其应用领域术语

2.1.1

大数据 big data

具有体量巨大、来源多样、生成极快、且多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据。

注：国际上，大数据的4个特征普遍不加修饰地直接用 volume、variety、velocity 和 variability 予以表述，并分别赋予了它们在大数据语境下的定义：

- a) 体量 volume: 构成大数据的数据集的规模。
- b) 多样性 variety: 数据可能来自多个数据仓库、数据领域或多种数据类型。
- c) 速度 velocity: 单位时间的数据流量。
- d) 多变性 variability: 大数据其他特征，即体量、速度和多样性等特征都处于多变状态。

2.1.2

数据生存周期 data lifecycle

将原始数据转化为可用于行动的知识的一组过程。

2.1.3

大数据参考体系结构 big data reference architecture

一种用作工具以便于对大数据内在的要求、设计结构和运行进行开放性探讨的高层概念模型。

注：比较普遍认同的大数据参考体系结构一般包含系统协调者、数据提供者、大数据应用提供者、大数据框架提供者和数据消费者等5个逻辑功能构件。

2.1.4

系统协调者 system orchestrator

大数据参考体系结构中的一种逻辑功能构件，它定义所需的数据应用活动并将它们整合到可运行的垂直系统中。

注1：系统协调者可以是人、软件或这二者。

注2：系统协调者一般包括：业务领导者、咨询师、数据科学家、信息体系结构设计师、软件体系结构设计师、安全体系结构设计师、个人信息保护体系结构设计师和网络体系结构设计师。

2.1.5

数据提供者 data provider

大数据参考体系结构中的一种逻辑功能构件，它将新的数据或信息引入大数据系统。

注：数据提供者一般包括：企业、公共机构、科学家、调研人员、从事数据搜索的工程师、网络应用软件、网络运营商和终端用户。