



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、网络运营商和末端用户。

2.1.6

大数据应用提供者 big data application provider

大数据参考体系结构中的一种逻辑功能构件,它执行数据生存周期操作,以满足系统协调者定义的需求以及安全和隐私保护需求。

注:大数据应用提供者一般包括:应用领域专家、平台领域专家和咨询师。

2.1.7

大数据框架提供者 big data framework provider

大数据参考体系结构中的一种逻辑功能构件,它建立一种计算框架,在此框架中执行转换应用,同时保护数据完整性和隐私。

注:大数据框架提供者一般包括:内嵌数据集集群、数据中心和云提供者。

2.1.8

数据消费者 data consumer

大数据参考体系结构中的一种逻辑功能构件,它是使用大数据应用提供者提供的应用的末端用户或其他系统。

注:数据消费者一般包括:末端用户、调研人员、应用和系统。

2.1.9

基础设施框架 infrastructure framework

由网络、计算、存储和环境等功能构件构成的一种集合。

注1:网络、计算、存储和环境的解释如下:

- a) 网络:支持将数据从一个资源传输到另一个资源的资源(如,已定义的物理资源、软件资源、虚拟资源等)。
- b) 计算:执行和驻留其他大数据系统构件(如,物理资源、操作系统、虚拟实现、逻辑分布)的软件的物理处理器和存储器。
- c) 存储:在大数据系统中保存数据的资源(如,存储器、本地磁盘、独立磁盘的软/硬件冗余阵列、存储域网、附网存储)。
- d) 环境:建立大数据系统时必须考虑的物理辅助资源(如,供电、冷却等)。

注2:这是大数据框架提供者可能提供的一种框架。

2.1.10

数据平台框架 data platform framework

用于指导实现结合相关应用程序接口(API)访问的逻辑数据组织和分发的集合。

注1:此类框架一般还包含数据注册和连同语义数据描述(如格式化本体或分类)的元数据服务。逻辑数据组织的覆盖范围从简单限定的平面文件到完全分布式关系数据存储或分栏数据存储。

注2:这是大数据框架提供者可能提供的一种框架。

2.1.11

处理框架 processing framework

覆盖为支持大数据应用实现所需基础设施软件的、定义数据的计算和处理的集合。

注:这是大数据框架提供者可能提供的一种框架。

2.1.12

消息/通信框架 messaging/communications framework

源于高性能计算环境、为水平扩展集群中节点之间的可靠查询、传输和接收数据提供API的集合。

注:这是大数据框架提供者可能提供的一种框架。

2.1.13

资源管理框架 resource management framework

大数据框架提供者可能提供的、利用数据本地化作为一种输入变量来确定是否安装新的处理框架

元素(如,主节点、处理节点、作业位置),从而实现对 CPU 和存储两大资源高效且有效管理的集合。

注:这是大数据框架提供者可能提供的一种框架。

2.1.14

大数据系统 big data system

实现大数据参考体系结构的全部或部分功能的系统。

2.1.15

大数据服务 big data service

基于大数据参考体系结构提供的数据服务。

2.1.16

垂直扩展 vertical scaling

为提高性能而提高处理速度、存储和内存等系统参数的过程。

2.1.17

水平扩展 horizontal scaling

将集成的一群个体资源作为一个单系统使用的过程。

2.1.18

大数据范例 big data paradigm

一种由水平耦合分布式数据系统和独立资源组成的、用于实现为有效处理众多数据集所必需的可伸缩性的知识。

2.1.19

大数据工程化 big data engineering

为适应大数据对于有效存储、操作和分析的需求而运用治理独立资源的先进技术构建可伸缩数据系统的过程。

2.1.20

大规模并行处理 massively parallel processing

多个处理器并行工作以执行一个特定计算任务的过程。

2.1.21

分布式文件系统 distributed file system

多个结构化数据集分布在一个或多个服务器集群的各个计算节点的文件系统。

注:此类系统中,数据可能分布在文件和/或数据集层,更为普遍的是在数据块这个层级分布,同时支持集群中多个节点与大型文件和/或数据集的不同部分交互。

2.1.22

分布式计算 distributed computing



一种覆盖存储层和处理层的、用于实现多类型程序设计算法模型的计算模式。

注:分布式计算结果通常加载到分析环境。MapReduce 是数据分布式计算中默认的处理构件。

2.1.23

分散-聚集 scatter-gather

大数据集的处理形式,其中所需的计算被划分并分布在集群的多个节点上,整体结果由每个节点的结果合并而成。

注:分散-聚集通常要求对处理软件的算法进行改变。

示例:MapReduce(包含 Map 和 Reduce 两个计算过程的一种计算模型)就是采用分散-聚集的处理形式。

2.1.24

流数据 streaming data

经由接口传递,从连续运行的数据源产生的数据。

2.1.25

非结构化数据 unstructured data

不具有预定义模型或未以预定义方式组织的数据。

2.1.26

大数据生存周期模型 lifecycle model for big data

用于描述大数据的“数据—信息—知识—价值”生存周期和指导大数据相关活动的模型；这些活动主要由收集、准备、分析和行动等阶段覆盖。

注：几个阶段的主要活动如下：

- a) 收集阶段：采集原始数据并按原始数据形式存储；
- b) 准备阶段：将原始数据转化为干净的、有组织的信息；
- c) 分析阶段：利用有组织的信息产生合成的知识；
- d) 行动阶段：运用合成的知识为组织生成价值。

2.1.27

读时模式 schema-on-read

一种数据模式应用；按此应用，在从数据库读取数据之前，先经过诸如转换、净化、整合之类准备步骤。

2.1.28

计算可移植性 computational portability

使计算移动到数据所在位置的能力。

2.1.29

真实性 veracity

数据在跨边界传送的情况下，与数据完整性和隐私保护相关的一种数据特征；亦简单指数据的准确性。

2.1.30

价值 value

从分析学角度考虑的数据对组织的重要性。

注：大数据应用领域越来越看重大数据带来的价值，确定数据的价值也趋向于作为大数据分析的一个重要目标。

2.1.31

波动性 volatility

数据结构随时间变化的趋势。

注：这个术语与大数据主要特征之一的“多变性(variability)”不同，多变性主要用于表述大数据的体量、速度和多样性等特征呈现的多变性。

2.1.32

正确性 validity

就数据预期用途而言的数据适当性。

2.1.33

大数据动态应用 big data velocity application

数据的收集、准备和分析(预警)在动态改变中发生，并可能在数据存储之前进行归纳或聚合。

2.1.34

大数据卷系统 big data volume system

在数据准备阶段前以数据原始形式存储的一种数据系统。

注：在这种系统中，在数据读出时开始启动准备阶段，因此被称为“读时模式”。

2.1.35

数据仓库 data warehouse

在数据准备之后用于永久性存储数据的数据库。

2.1.36

动态数据 data in motion

处于活动状态,其典型特征表现为大数据的速度和多变性特征的数据。

注:它们在网络上传输或暂时驻留于计算机内存中供读取或更新。对它们以实时或近实时方式进行处理和分析。

2.1.37

静态数据 data at rest

处于静止状态,其典型特征表现为大数据的体量和多样性特征的数据。

注:它们通常是存储于物理媒体中的数据。

2.1.38

非关系模型 non-relational models

用于数据存储和处理、不遵循关系代数的逻辑数据模型。

注:非关系模型也常称为 NoSQL,通常理解为非 SQL(结构化查询语言)或不仅是 SQL。

2.1.39

联合数据库系统 federated database system

一种元数据库管理系统,它透明地将多个自治数据库系统映射到一个单一联合数据库。

2.1.40

数据科学 data science

根据原始数据,经过整个数据生存周期过程凭借经验合成可用于行动的知识的一种科学。

2.1.41

数据科学范例 data science paradigm

通过发现、假设和假设测试过程直接从数据萃取的可用于行动的知识。

2.1.42

数据科学家 data scientist

数据科学专业人员:他们具有足够的业务需求管理机制方面的知识、领域知识、分析技能以及用于管理数据生存周期中每个阶段的端到端数据过程的软件和系统工程知识。

2.1.43

数据治理 data governance

对数据进行处置、格式化和规范化的过程。

注 1:数据治理是数据和数据系统管理的基本要素。

注 2:数据治理涉及数据全生存周期管理,无论数据是处于静态、动态、未完成状态还是交易状态。

2.1.44

开放数据 open data

可为其他数据使用的数据。

2.1.45

链接数据 linked data

连接其他数据的数据。

2.1.46

数据集 data set

数据记录汇聚的数据形式。

注:它可以具有大数据的体量、速度、多样性和易变性特征。数据集的特征表征的是数据本身或静态数据,而数据的特征,当其在网络上传输时或暂时驻留于计算机存储器中以备读出或更新时,表征的是动态数据。

2.1.47

追溯 provenance

对数据集的历史元数据的讨论。

注 1：此词条的中文名是对同一个英文名词的动词性定义表示。

注 2：这是大数据分析中的一个必不可少的因素。

2.1.48

分析 analytics

根据信息合成知识的过程。

2.1.49

分析过程特征 analytic processes characteristics

用以表征大数据分析过程的发现、开发和应用。

“发现”是形成最初的假设性构想，“开发”是针对具体构想构建分析过程，“应用”则是将分析结果打包到特定运行的系统。

2.1.50

共享磁盘文件系统 shared-disk file systems

使用单一存储池且与多个计算资源关联的存储数据的方法。

注：此类系统的技术实现支持同时从多个节点访问许多大型数据集。

示例 1：

存储域网(Storage Area Networks, 简称 SAN)。

示例 2：

附网存储(Network Attached Storage, 简称 NAS)。

2.1.51

数据特征层次 data characteristic hierarchy

从不同粗细粒度表征数据特征的数据层次结构。

注：大数据的特征层级一般包括以下层级：

- 数据元素；
- 记录(数据元素的汇集)；
- 数据集(记录的汇集)；
- 多数据集(数据集的汇集)。

2.1.52

可伸缩流处理 scalable stream processing

数据存储器之间动态数据的处理形式。

注：主要用于数据的过滤、转换或路由选择。对于大数据流，流处理往往是可伸缩的，以便支持分布式处理和流水线式处理。

2.1.53

可伸缩数据存储 scalable data stores

用以支持应对数据存储无限增长的一种存储技术。

注：此类技术的运用往往伴之以容错能力，以便应对大数据系统构件的某些失效。

2.1.54

本体 ontology

在大数据语境下，它是一些约束后续各种不同层次逻辑模型的语义模型。

注：本体，从本质上看，既可以是概括性的，也可以是极其专门化的。

2.1.55

分类 taxonomies

在数据分析语境下表示关于数据元素关系的元数据。

注：它是实体之间的层次关系，在这种情况下，一个数据元素被分解为较小的组成部分。

2.1.56

图解模型 graphical model

一种可以呈现数据元素之间关系的大数据记录存储类型。

注：在此种模型中，数据元素是节点，关系表现为节点之间链接。

2.1.57

复杂性 complexity

在大数据语境下，复杂性是指各数据元素之间内部关系的或跨数据记录之间的一种互动关系繁简程度。

2.1.58

资源协商 resource negotiation

一种支持多租户以及要求高可用性和低延迟的环境的资源访问模式。

注：按此模式，资源管理器是若干节点管理器的集线器；各个客户（或用户）依次请求节点管理器中的应用管理器，紧接前一个请求者的后一个请求者分配到同一个或不同的节点管理器的应用管理器。根据中央处理器（CPU）和存储器可用情况为所请求的任务确定先后次序并在节点提供适当的处理资源。

2.1.59

集群管理 cluster management

在以非关系模型方式驻留数据的集群资源之间提供通信的一种机制。

2.1.60

软件定义存储 software defined storage

利用软件来确定存储器的动态层级分配的一种存储管理技术。

注：此类技术可以以较低的存储开销来维护必要的数据库检索性能，常用于内存、高速缓存、固态硬盘、网络驱动等应用领域。

2.1.61

软件定义网络 software defined network; SDN

支持作为大数据关键实现的网络资源高效且有效管理的一种技术。

注：亦称虚拟网络（virtual network），与传统的用于数据、管理、I/O（输入/输出）和控制的专用物理网络链接不同，SDN 包含资源池化的链接和实际交换设施，实施特定功能和特定应用的按需分配（包括传输的原始带宽、服务质量和数据路由等）。

2.1.62

网络功能虚拟化 network function virtualization

对路由器/路由选择、周界防护、远程访问鉴别以及网络流量/载荷监控等网络功能的虚拟应用实现。

注：网络功能虚拟化支持信息系统的高弹性、容错和资源管理，是应对大数据巨大数据体量下用户数据连接的峰、谷起伏问题的至关重要的应用。

2.1.63

本地虚拟化 native virtualization

大数据环境下的一种虚拟化基本形式，按此种形式，在本地裸机上运行管理程序，该程序管理由操作系统和应用组成的多个虚拟机。

2.1.64

主机虚拟化 hosted virtualization

大数据环境下的一种虚拟化基本形式,按此种形式,在本地裸机上运行操作系统,在驻留客户操作系统和应用的顶层运行管理程序。

2.1.65

容器式虚拟化 containerized virtualization

大数据环境下的一种虚拟化基本形式,按此种形式,在本地裸机上运行操作系统,管理程序功能嵌入操作系统中,应用在内部各个容器中运行,控制或限制对操作系统和物理机器资源的访问。

2.2 密切相关的通用术语

2.2.1

数据 data

信息的可再解释的形式化表示,以适用于通信、解释或处理。

注:可以通过人工或自动手段处理数据。

[GB/T 5271.1—2000,定义 01.01.02]

2.2.2

数据处理 data processing

数据操作的系统执行。

注:术语“数据处理”不能用作“信息处理”的同义词。

[GB/T 5271.1—2000,定义 01.01.06]

2.2.3

数据管理 data management

在数据处理系统中,提供对数据的访问,执行或监视数据的存储,以及控制输入输出操作等功能。

[GB/T 5271.1—2000,定义 01.08.02]

2.2.4

关系模型 relational model

结构基于一组关系的数据模型。

示例:结构查询语言(SQL)即表示这样一种模型。

[GB/T 5271.17—2010,定义 17.04.04]

2.2.5

关系数据库 relational database

数据按关系模型来组织的数据库。

[GB/T 5271.17—2010,定义 17.04.05]

2.2.6

分布式数据处理 distributed data processing; DDP

将操作分散到计算机网络的各结点进行的数据处理。

注:DDP需要借助各结点之间的数据通信做到集体协作。

[GB/T 5271.18—2008,定义 18.01.08]

2.2.7

元数据 metadata

关于数据或数据元素的数据(可能包括其数据描述),以及关于数据拥有权、存取路径、访问权和数据易变性的数据。

[GB/T 5271.17—2010,定义 17.06.05]

2.2.8

语义元数据 semantic metadata

元数据的一个类型;给出有助于恰当理解数据元素的定义性描述的元数据,或,作为数据元素的定义的一类元数据。

2.2.9

数据挖掘 data mining

从大量的数据中通过算法搜索隐藏于其中信息的过程。

注:一般通过包括统计、在线分析处理、情报检索、机器学习、专家系统(依靠过去的经验法则)和模式识别等方法来实现。

[GB/T 33745—2017,定义 2.5.3]

2.2.10

云计算 cloud computing

一种通过网络将可伸缩、弹性的共享物理和虚拟资源池以按需自服务的方式供应和管理的模式。

注:资源包括服务器、操作系统、网络、软件、应用和存储设备等。

[GB/T 32400—2015,定义 3.2. 5]

2.2.11

物联网 internet of things; IOT

通过感知设备,按照约定协议,连接物、人、系统和信息资源,实现对物理和虚拟世界的信息进行处理并作出反应的智能服务系统。

注:物即物理实体。

[GB/T 33745—2017,定义 2.1.1]

2.2.12

数据中心 data center

由计算机场站(机房)、机房基础设施、信息系统硬件(物理和虚拟资源)、信息系统软件、信息资源(数据)和人员以及相应的规章制度组成的组织。

[GB/T 33136—2016,定义 3.1.1]

2.2.13

结构化数据 structured data

一种数据表示形式,按此种形式,由数据元素汇集而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。

2.2.14

表结构 table structure

为主体层内容提供表示语义的一种存储范例。

2.2.15

键值结构 key-value structure

一种存储范例,按此范例存储的记录由值中的关键词和一串数据组成。

注:在键值结构下,数据通过关键词检索,同时由非关系数据库软件处理对于值中的数据的访问。其作用相当于带单一索引字段和列的关系数据库表的子集和/或简化版。其变体之一是文档存储,在这种情况下,文档有多个值字段,其中任何一个都可以用作索引/关键词。

参 考 文 献

- [1] GB/T 5271.1—2000 信息技术 词汇 第 1 部分:基本术语
- [2] GB/T 5271.17—2010 信息技术 词汇 第 17 部分:数据库
- [3] GB/T 5271.18—2008 信息技术 词汇 第 18 部分:分布式数据处理
- [4] GB/T 32400—2015 信息技术 云计算 概览与词汇
- [5] GB/T 33136—2016 信息技术服务 数据中心服务能力成熟度模型
- [6] GB/T 33745—2017 物联网 术语



索引

汉语拼音索引

B	
本地虚拟化	2.1.63
本体	2.1.54
表结构	2.2.14
波动性	2.1.31
C	
处理框架	2.1.11
垂直扩展	2.1.16
D	
大规模并行处理	2.1.20
大数据	2.1.1
大数据参考体系结构	2.1.3
大数据动态应用	2.1.33
大数据范例	2.1.18
大数据服务	2.1.15
大数据工程化	2.1.19
大数据卷系统	2.1.34
大数据框架提供者	2.1.7
大数据生存周期模型	2.1.26
大数据系统	2.1.14
大数据应用提供者	2.1.6
动态数据	2.1.36
读时模式	2.1.27
F	
非关系模型	2.1.38
非结构化数据	2.1.25
分布式计算	2.1.22
分布式数据处理	2.2.6
分布式文件系统	2.1.21
分类	2.1.55
分散-聚集	2.1.23
分析	2.1.48
分析过程特征	2.1.49
复杂性	2.1.57

G	
共享磁盘文件系统	2.1.50
关系模型	2.2.4
关系数据库	2.2.5
J	
基础设施框架	2.1.9
集群管理	2.1.59
计算可移植性	2.1.28
价值	2.1.30
键值结构	2.2.15
结构化数据	2.2.13
静态数据	2.1.37
K	
开放数据	2.1.44
可伸缩流处理	2.1.52
可伸缩数据存储	2.1.53
L	
联合数据库系统	2.1.39
链接数据	2.1.45
流数据	2.1.24
R	
容器式虚拟化	2.1.65
软件定义存储	2.1.60
软件定义网络	2.1.61
S	
数据	2.2.1
数据仓库	2.1.35
数据处理	2.2.2
数据管理	2.2.3
数据集	2.1.46
数据科学	2.1.40
数据科学范例	2.1.41

数据科学家	2.1.42
数据平台框架	2.1.10
数据生存周期	2.1.2
数据特征层次	2.1.51
数据提供者	2.1.5
数据挖掘	2.2.9
数据消费者	2.1.8
数据治理	2.1.43
数据中心	2.2.12
水平扩展	2.1.17



T

图解模型	2.1.56
------------	--------

W

网络功能虚拟化	2.1.62
物联网	2.2.11

X

系统协调者	2.1.4
消息/通信框架	2.1.12

Y

语义元数据	2.2.8
元数据	2.2.7
云计算	2.2.10

Z

真实性	2.1.29
正确性	2.1.32
主机虚拟化	2.1.64
追溯	2.1.47
资源管理框架	2.1.13
资源协商	2.1.58

英文对应词索引

A

analytic processes characteristics	2.1.49
analytics	2.1.48

B

big data	2.1.1
big data application provider	2.1.6
big data engineering	2.1.19
big data framework provider	2.1.7
big data paradigm	2.1.18
big data reference architecture	2.1.3
big data service	2.1.15
big data system	2.1.14
big data velocity application	2.1.33
big data volume system	2.1.34

C

cloud computing	2.2.10
cluster management	2.1.59
complexity	2.1.57
computational portability	2.1.28

containerized virtualization	2.1.65
------------------------------------	--------

D

data	2.2.1
data at rest	2.1.37
data center	2.2.12
data characteristic hierarchy	2.1.51
data consumer	2.1.8
data governance	2.1.43
data in motion	2.1.36
data lifecycle	2.1.2
data management	2.2.3
data mining	2.2.9
data platform framework	2.1.10
data processing	2.2.2
data provider	2.1.5
data science	2.1.40
data science paradigm	2.1.41
data scientist	2.1.42
data set	2.1.46
data warehouse	2.1.35
DDP	2.2.6
distributed computing	2.1.22
distributed data processing	2.2.6
distributed file system	2.1.21

F

federated database system	2.1.39
---------------------------------	--------

G

graphical model	2.1.56
-----------------------	--------

H

horizontal scaling	2.1.17
hosted virtualization	2.1.64

I

infrastructure framework	2.1.9
internet of things	2.2.11
IOT	2.2.11

K

key-value structure	2.2.15
---------------------------	--------

L

lifecycle model for big data	2.1.26
linked data	2.1.45

M

massively parallel processing	2.1.20
messaging /communications framework	2.1.12
metadata	2.2.7

N

native virtualization	2.1.63
network function virtualization	2.1.62
non-relational models	2.1.38

O

ontology	2.1.54
open data	2.1.44

P

processing framework	2.1.11
provenance	2.1.47

R

relational database	2.2.5
relational model	2.2.4
resource management framework	2.1.13
resource negotiation	2.1.58

S

scalable data stores	2.1.53
scalable stream processing	2.1.52
scatter-gather	2.1.23
schema-on-read	2.1.27
SDN	2.1.61
semantic metadata	2.2.8
shared-disk file systems	2.1.50
software defined network	2.1.61
software defined storage	2.1.60
streaming data	2.1.24
structured data	2.2.13
system orchestrator	2.1.4

T

table structure 2.2.14

taxonomies 2.1.55

U

unstructured data 2.1.25

V

validity 2.1.32

value 2.1.30

veracity 2.1.29

vertical scaling 2.1.16

volatility 2.1.31
