

中华人民共和国金融行业标准

JR/T 0236—2021

---

金融大数据 术语

Financial big data—Terminologies

行业标准信息服务平台

2021 - 12 - 29 发布

2021 - 12 - 29 实施

---

中国人民银行 发布

## 目 次

前言.....	II
引言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 基础概念.....	1
4 参考体系和生命周期.....	6
5 治理与管理.....	11
6 支撑运行环境.....	15
附录（资料性）本文件中术语描述的策略.....	19
参考文献.....	20

行业标准信息服务平台

## 前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

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专利的责任。

本文件由中国人民银行提出。

本文件由全国金融标准化技术委员会（SAC/TC 180）归口。

本文件起草单位：中国人民银行科技司、成方金融信息技术服务有限公司、中国农业银行股份有限公司、中国人民银行金融信息中心、中国建设银行股份有限公司、中国人民银行重庆营业管理部、中证信息技术服务有限责任公司、华为技术有限公司、成方金融科技有限公司、中国工商银行股份有限公司、海通证券股份有限公司、工业和信息化部电子第五研究所、北京国家金融标准化研究院有限责任公司。

本文件主要起草人：李伟、陈立吾、杨富玉、周祥昆、虞瑾、齐小东、何军、梁静、丁俨、时向一、郝洁、李宽、王鹏、艾霁坤、聂晓峰、徐光贤、朱建强、孟桂清、刘静芳、曹建勇、刘超、刘启滨、秦逞、李向东、冯晋雯、赵华、符海芳、文州、汪星辰、沈云明、吴保杰、刘舒婷、柴思跃、谢彦丽、李琪。

行业标准信息服务平台

## 引 言

随着大数据技术在金融领域的广泛深入应用，大数据技术已经成为促进金融科技创新，驱动监管科技智能化的重要技术支撑，在规范社会治理、提供金融服务等方面发挥着重要作用。在金融应用场景下，一方面，大数据术语可能会被赋予特定的金融含义，部分术语从信息技术视角看不属于狭义的大数据概念，需要进一步定义，供金融行业参考；另一方面，在不同的场合，对大数据的概念存在不同的表述，导致在大数据管理与使用等过程中对大数据的实质性内容认知不同，可能影响信息传导和业务需求的表达。基于以上原因，有必要尽快统一金融领域大数据相关概念，促进快速形成共识。

本文件在收集现有国家标准、行业标准以及国际标准中大数据相关术语的基础上，通过分析归纳，结合我国金融应用特点，形成了金融大数据的术语标准。

本文件通过描述金融大数据的基础概念、参考体系和生命周期、治理与管理、支撑运行环境等方面的常用术语，以期望减少相关人员在沟通中的分歧，促进大数据在金融行业中的应用。

行业标准信息服务平台

# 金融大数据 术语

## 1 范围

本文件界定了金融大数据领域中的常用术语。

本文件适用于在金融领域中涉及到的与大数据相关的信息沟通。

注：本文件中术语描述的策略见附录。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 5271—2021 信息技术 词汇

GB/T 25000—2021 系统与软件工程 系统与软件质量要求和评价（SQuaRE）

GB/T 35274—2017 信息安全技术 大数据服务安全能力要求

GB/T 35295—2017 信息技术 大数据 术语

GB/T 34960—2018 信息技术服务 治理

GB/T 36073—2018 数据管理能力成熟度评估模型

ISO/IEC 2382:2015 Information technology — Vocabulary

ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary

ISO/IEC 20547:2020 Information technology — Big data reference architecture

ISO/IEC 20944:2013 Information technology — Metadata Registries Interoperability and Bindings (MDR-IB)

ISO/IEC 38505:2018 Information technology — Governance of IT — Governance of data

ISO 55000:2014 Asset management — Overview, principles and terminology

## 3 基础概念

### 3.1

#### 大数据 big data

海量的数据集，其数据在本质上具有体量大、种类多、变化快、变数多的特征，需要一种易扩展的技术来有效存储、处理、管理和分析。

注：1. 大数据通常以多种不同方式使用，例如，作为某种用于处理大数据海量数据集的易扩展技术的名称。

2. GB/T 35295—2017《信息技术 大数据 术语》中 2.1.1 对大数据的定义为“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”。

3. 在某些非工程性研讨的场合，“大数据”一词的外延可能被扩大到所有的数据。

[来源：ISO/IEC 20546:2019, 3.1.2, 有修改]

### 3.2

**金融大数据 financial big data**

金融领域中的大数据。

注：在某些非工程性研讨的场合，“金融大数据”一词的外延可能被扩大到所有的金融数据，也可能被限定在某一特殊的领域，例如银行大数据、证券大数据、保险大数据等。

3.3

**信息 information**

在信息处理中，关于客体（如事实、事件、事物、过程或想法，包括概念）的知识，这些客体在一定的场合中具有特定的含义。

注：在某些非工程性研讨的场合，“信息”和“数据”可能没有被严格地区分。

[来源：GB/T 5271.1—2000，01.01.01，有修改]

3.4

**数据 data**

任何以电子或者其他方式对信息的记录。

注：1. 可以通过人工或自动手段处理数据。

2. 在某些非工程性研讨的场合，“数据”有可能被认为是“大数据”的一个子集。

3. ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary 中 3.1.5 对数据的定义与此等同。

4. GB/T 5271.1—2000《信息技术 词汇 基本术语》中 01.01.02 对数据的定义为“信息的可再解释的形式化表示，以适用于通信、解释或处理”。

3.5

**证券交易数据 securities trading data**

在证券交易过程中产生的数据。

3.6

**证券期货行情数据 securities and futures market data**

有关证券期货市场价格的数据。

3.7

**数据隐私 data privacy**

记录与描述自然人的私人生活安宁和不愿为他人知晓的私密空间、私密活动、私密信息的数据。

注：在不同工作场合、不同业务场景和不同数据组合下，“数据隐私”可能对应不同的外延。

3.8

**目标实体 target entity**

与用户相关的基础事物，关于这些事物的信息应保留并被测量。

注：1. 目标实体同义词可能包括信息产品输入和工作成果。

2. 目标实体的示例有：体系结构、上下文模式、概念模型、逻辑模型、物理数据模型、数据字典、文档、数据文件、数据库管理、关系数据库管理系统、表单、显现设备等。

3. 目标实体由性质精确定义。性质的示例有：属性、元素、信息、元数据、词汇表、数据格式、数据项、数据值、信息项、信息项内容、数据记录等。

[来源：GB/T 25000.24—2017，4.36，有修改]

3.9

**属性 attribute**

目标实体的固有性质或特性，能用人工或自动手段定量或定性地区分。

[来源：GB/T 25000.24—2017，4.2]

3.10

**数据变异性 data variability**

数据集传输速率、格式或结构、语义或质量的变化。

[来源：ISO/IEC 20546:2019, 3.1.13]

## 3.11

**数据多样性 data variety**

数据集的各种格式、逻辑模型、时间尺度和语义。

注：数据多样性指不规则数据结构或异构数据结构，及导航、查询和数据类型划分。

[来源：ISO/IEC 20546:2019, 3.1.14]

## 3.12

**数据速度 data velocity**

创建、传输、存储、分析或可视化数据的流速。

[来源：ISO/IEC 20546:2019, 3.1.15]

## 3.13

**数据真实性 data veracity**

数据的完整性或准确性。

注：数据真实性指用以支持实时决策的描述性数据和对于对象的自查询。

[来源：ISO/IEC 20546:2019, 3.1.16]

## 3.14

**数据波动性 data volatility**

与数据的一段时间变化率有关的特征。

[来源：ISO/IEC 20546:2019, 3.1.17]

## 3.15

**数据体量 data volume**

数据量的范围，与数据处理过程中的计算和存储资源及对这些资源的管理有关。

注：数据体量在处理大型数据集中变得非常重要。

[来源：ISO/IEC 20546:2019, 3.1.18]

## 3.16

**数据模型 data model**

根据数据在信息系统的正式描述和所使用的数据库管理系统的要求确定的在数据库中构建数据的模式。

注：大数据不需要数据模型的观点是不正确的，至少是不完整的。大数据的数据模型往往在对大数据进行处理时，根据大数据的特征和对大数据加工的目的构建。

[来源：ISO/IEC 20546:2019, 3.1.8, 有修改]

## 3.17

**上下文模式 contextual schema**

应用了数据模型的使用周境边界的正规描述。

注：这是对业务信息需求的高层描述，上下文模式比概念模型更一般化，包括体系结构的（系统）周境的整体视图。

[来源：GB/T 25000.24—2017, 4.4, 有修改]

## 3.18

**数据集 data set (dataset)**

可用一种或多种格式访问或下载的可标识的数据集合。

注：数据集中的数据可能是可唯一定位但非为某一特定目的组织的，也可能是非结构化的。

[来源：ISO/IEC 20546:2019, 3.1.11, 有修改]

3.19

**数据项 data item**

在特定上下文内数据的最小可识别单位，其定义、标识、允许值和其他信息由一组属性指定。

注：1. 字段被认为是数据项的同义词。

2. 数据项是数据值的物理对象“容器”。

[来源：GB/T 25000.24—2017，4.9]

3.20

**数据记录 data record**

一个单元化的相关数据项的集合。

[来源：GB/T 25000.24—2017，4.15]

3.21

**数据文件 data file**

一个单元化的相关数据记录的集合。

[来源：GB/T 25000.24—2017，4.7]

3.22

**主数据 master data**

组织中需要跨系统、跨部门共享的核心业务实体数据。

注：大数据意义下的主数据可能是原始数据中的，也可能是加工后的数据中的，两者并不等同。

[来源：GB/T 36073—2018，3.12，有修改]

3.23

**数据字典 data dictionary**

收集有关数据的信息，如名称、描述、创建者、所有者、出处、不同语言的翻译以及用法。

[来源：GB/T 25000.24—2017，4.6]

3.24

**数据格式 data format**

对于数据的储存或者显现的管理安排。

注：数据格式可以参照数据类型以及数据项的长度。

[来源：GB/T 25000.24—2017，4.8，有修改]

3.25

**数据类型 data type (datatype)**

指定数据结构和一组允许操作的一组数据对象，以便这些数据对象在执行其中任何一个操作时作为操作数。

注：ISO/IEC 2382:2015 Information technology — Vocabulary中17.6.6对数据类型的定义为“以这些值的属性和对这些值的操作为特征的不同值集”。

[来源：ISO/IEC 20546:2019，3.1.12，有修改]

3.26

**数据值 data value**

数据项的内容。

注：1. GB/T 25000.12—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE）》的5.1.1中说明，从“固有”的角度看，数据质量是指数据本身，如数据域值和可能的限制。

2. 数据值是对目标实体的属性赋予的数值或类别。

[来源：GB/T 25000.24—2017，4.17，有修改]

3.27

**数据指纹 digital fingerprint**

部署算法的技术，该技术通过分析设备上的大量技术特征和设置，生成独特的标识符，该标识符可对产生机器识别码的特定计算设备进行识别，并可对个人身份进行识别。

[来源：ISO 19731:2017, 3.17]

## 3.28

**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进行有效描述。

注：一般来说，原始未经处理的大数据不是结构化数据，但处理的结果往往采用结构化的方式组织。

[来源：GB/T 35295—2017, 2.2.13, 有修改]

## 3.29

**半结构化数据 semi-structured data**

包含用来分隔语义元素以及对记录和字段进行分层的标记，但不符合关系模型的数据。

示例：可扩展标记语言(Extensible Markup Language, XML)和Java脚本对象表示法(JavaScript Object Notation, JSON)都是半结构化数据的例子。

注：1. 半结构化数据往往不完全符合除关系模型外其他形式的关联。

2. 半结构化数据的一般特点是属于同一类的实体可以有不同的属性。

3. 半结构化数据往往也被称为自描述的数据，其顺序往往并不重要。

4. ISO/IEC 20944-1:2013 Information technology — Metadata Registries Interoperability and Bindings (MDR-IB)中3.21.12.21对半结构化数据的定义为“组件的数据类型及其标签未预先确定的聚合数据类型”。

## 3.30

**非结构化数据 unstructured data**

特征为除了记录或文件级别外没有任何结构的数据。

示例：自由文本是非结构化数据的一个例子。

注：总体而言，非结构化数据不是由数据元素组成。

[来源：ISO/IEC 20546:2019, 3.1.37, 有修改]

## 3.31

**静态数据 data at rest**

处于静止状态，具有大数据的体量大和多样性特征的数据。

示例：静态数据可能不是大数据，例如应用系统的配置数据。

注：静态数据通常是存储于物理媒体中的数据。

[来源：GB/T 35295—2017, 2.1.37, 有修改]

## 3.32

**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。

示例：动态数据可能不是大数据，例如应用系统在某个时点的交易量。

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

[来源：GB/T 35295—2017, 2.1.36, 有修改]

## 3.33

**开放数据 open data**

可被其他数据使用的数据。

注：开放数据有可能有一个开放的范围，例如客户信息在核心银行系统内是对交易系统开放的，但并不对核心银行

系统外的其他系统开放。

[来源：GB/T 35295—2017，2.1.44，有修改]

### 3.34

**链接数据 linked data**

连接其他数据的数据。

[来源：GB/T 35295—2017，2.1.45]

### 3.35

**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。

[来源：GB/T 35295—2017，2.1.24]

### 3.36

**关系模型 relational model**

结构基于一组关系的数据模型。

[来源：GB/T 5271.17—2010，17.04.04，有修改]

### 3.37

**非关系模型 non-relational model**

在存储和处理数据时不遵从关系模型的逻辑数据模型。

[来源：ISO/IEC 20546:2019，3.1.28]

### 3.38

**内源性数据 endogenous data**

产生规则由系统协调者确定，并可在需要时由系统协调者协调数据提供者、大数据应用提供者、大数据框架提供者产生、变更、废止使用的数据。

[来源：GB/T 31186.1—2014，3.7，有修改]

### 3.39

**外源性数据 exogenous data**

在系统协调者、大数据应用提供者、大数据框架提供者外部产生，产生规则可能公开也可能不公开，数据提供者在提供数据时不能确定其可用状态，可在系统协调者不知道的情况下产生、变更、废止使用的数据。

[来源：GB/T 31186.1—2014，3.8，有修改]

## 4 参考体系和生命周期

### 4.1

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

一种用作工具以便于对数据内在要求、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。

[来源：GB/T 35295—2017，2.1.3]

### 4.2

**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，可定义所需的数据应用活动并将其整合到可运行的垂直系统中。

示例：在金融行业中，系统协调者可能是金融监督管理部门或相关行业协会。在一个金融机构内部，系统协调者可

能是高级管理层或特别授权的大数据管理部门。

注：1. 系统协调者可以是人、软件或这二者。

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。

[来源：GB/T 35295—2017，2.1.4，有修改]

#### 4.3

##### 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，可将新的数据或信息引入大数据系统。

示例：在金融行业中，数据提供者可能是各金融机构。在一个金融机构内部，数据提供者可能是某个应用系统、某个业务部门或某个金融业务的客户。

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。

[来源：GB/T 35295—2017，2.1.5，有修改]

#### 4.4

##### 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，可执行数据生命周期操作，以满足系统协调者定义的需求以及安全和隐私保护的需求。

示例：在金融行业中，大数据应用提供者可能是金融监督管理机构、金融服务提供商或某些厂商。在一个金融机构内部，大数据应用提供者可能是信息科技管理部门、应用研发部门或相关业务部门。

注：大数据应用提供者一般包括：大数据应用领域专家、大数据平台领域专家和大数据咨询师。

[来源：GB/T 35295—2017，2.1.6，有修改]

#### 4.5

##### 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，可建立一种计算框架，在此框架中执行转化应用，同时保护数据完整性和隐私。

示例：在金融行业中，大数据框架提供者可能是金融基础设施。在一个金融机构内部，大数据框架提供者可能是数据中心。

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。

[来源：GB/T 35295—2017，2.1.7，有修改]

#### 4.6

##### 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，是使用大数据应用提供者所提供的应用的末端用户或其他系统。

示例：在金融行业中，数据消费者可能是各业务部门或各业务应用系统。

注：数据消费者一般包括：末端用户、调研人员、应用和系统。

[来源：GB/T 35295—2017，2.1.8，有修改]

#### 4.7

##### 大数据生命周期模型 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、分析和行动等阶段组成。

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。

[来源：GB/T 35295—2017，2.1.26，有修改]

#### 4.8

##### 数据生命周期 data lifecycle

数据从产生，经过数据采集、数据传输、数据存储、数据处理（如计算、分析、可视化等）、数据交换，直至数据销毁等各种生存形态的演变过程。

注：ISO/IEC 20547-3:2020 Information technology — Big data reference architecture的3.16将数据生命周期定义为“数据管理的各个阶段”。

[来源：GB/T 35274—2017，3.2，有修改]

#### 4.9

##### 大数据应用 big data application

执行数据生命周期相关的数据采集、数据传输、数据存储、数据处理（如计算、分析、可视化等）、数据交换、数据销毁等数据活动，在大数据平台运行，并提供大数据服务的各种应用系统。

注：1. 本术语定义中提及的数据采集、数据传输、数据存储、数据处理、数据交换、数据销毁等活动，分别是大数据采集、大数据传输、大数据存储、大数据处理、大数据交换、大数据销毁等活动的超集，在没有实质性区别时，可以视为同一概念。

2. 应用系统可以理解为软件与平台。

[来源：GB/T 35274—2017，3.5，有修改]

#### 4.10

##### 数据采集 data collection

通过不同方式汇集信息的过程。

[来源：ISO 19731:2017，3.14]

#### 4.11

##### 抽取—转换—加载 extract, transform and load; ETL

抽取、转换和装入。

[来源：ISO/TR 25100:2012，2.2.12，有修改]

#### 4.12

##### 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。

2. 这是大数据分析中的一个必不可少的因素。

[来源：GB/T 35295—2017，2.1.47]

#### 4.13

##### 大数据传输 big data transmission

将大数据由信源传输到信宿的过程。

注：1. 大数据传输的过程可能涉及到数据交换和数据同步。

2. 流数据的传输过程可能与数据处理和大数据消费过程交织。

3. 大数据传输可能需要事先明确传输的原则并通过必要的手段进行管控。

#### 4.14

##### 信源 message source

通信系统中视作信息来源的部分。

[来源：ISO/IEC 2382:2015，2123206]

## 4.15

**信宿 message sink**

通信系统中视作信息归宿的部分。

[来源: ISO/IEC 2382:2015, 2123207]

## 4.16

**数据交换 data interchange**

为满足不同平台或应用间对数据资源的传送和处理需要,依据一定的原则,采取相应的技术和工具,实现不同平台或应用间数据资源的流动过程。

[来源: GB/T 35274—2017, 3.11, 有修改]

## 4.17

**数据同步 data synchronization**

在业务交易过程中,为确保所有参与方的信息系统中相关记录信息的当前状态相同而持续调配记录信息的过程。

[来源: ISO/IEC 15944—8:2012, 3.35]

## 4.18

**非易失存储 nonvolatile storage**

在切断电源后数据不会丢失的存储设备。

[来源: ISO/IEC 2382:2015, 2125560]

## 4.19

**大数据存储 big data storage**

将大数据保存在非易失存储中的过程。

## 4.20

**数据处理 data processing**

数据操作的系统执行。

注: 1. 对数据执行合并、排序等算术或逻辑运算,或对文本执行编辑、排序、合并、存储、检索、显示、打印等操作都是数据处理。

2. 术语“数据处理”不能用作“信息处理”的同义词。

[来源: ISO/IEC 20546:2019, 3.1.9]

## 4.21

**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、整合等准备步骤。

[来源: GB/T 35295—2017, 2.1.27, 有修改]

## 4.22

**大数据脱敏 big data desensitization**

将大数据中涉及到的敏感信息进行匿名化、泛化、可逆化、随机化等处理的过程。

## 4.23

**大数据消费 big data consumption**

应用大数据处理的结果实现特定业务目的的过程。

注: 大数据消费的过程中,可能需要对大数据处理的结果再次进行数据处理,以实现特定的业务目的。

## 4.24

**价值 value**

从分析学角度考虑的数据对组织的重要性。

[来源：GB/T 35295—2017，2.1.30]

4.25

**数据共享 data sharing**

让不同的大数据用户能够访问大数据服务整合的各种数据资源，并通过大数据服务或数据交换技术对这些数据进行相关的计算、分析、可视化等处理。

[来源：GB/T 35274—2017，3.12]

4.26

**数据挖掘 data mining**

从大量数据中通过算法搜索隐藏于其中的信息的过程。

注：一般通过统计、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）和模式识别等方法实现。

[来源：GB/T 33745—2017，2.5.3，有修改]

4.27

**分析 analytics**

根据信息合成知识的过程。

[来源：GB/T 35295—2017，2.1.48]

4.28

**数据分析 data analytics**

由数据采集、数据验证、数据处理组成的复合概念，包括数据量化、数据可视化和数据解释。

注：1. 数据分析用于理解数据表示的对象，针对给定情况做出预测，并建议实现目标的步骤。从数据分析中获得的见解可用于各种目的，例如决策、研究、可持续发展、设计、规划等。

2. 本词条中“数据采集”对应 ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary 中的原文为“data acquisition, data collection”。

[来源：ISO/IEC 20546:2019，3.1.6，有修改]

4.29

**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”是将分析结果打包到特定运行的系统。

[来源：GB/T 35295—2017，2.1.49，有修改]

4.30

**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。

[来源：GB/T 35295—2017，2.1.15]

4.31

**大数据处置 big data disposal**

将大数据及其数据处理的中间结果进行不可恢复的、销毁的过程。

4.32

**数据清洗 data cleaning**

通过检测和纠正（或消除）数据中的缺陷和错误来改善数据质量的过程。

[来源：ISO 5127:2017，3.1.11.21]

4.33

**数据清理 data cleansing**

检查和纠正数据以确保数据符合标准格式的过程。

注：可能会对不完整、格式不正确、陈旧或重复的数据等进行纠正。通常在合并数据集或将数据从一个系统或数据库转换到另一个系统或数据库之前实施。

[来源：ISO 13008:2012, 3.6]

#### 4.34

##### 数据可视化 data visualization

以图形、图像、地图、动画等生动、易于理解的方式来展示数据，诠释数据之间的关系、趋势、规律，以更好地理解数据的方法。

#### 4.35

##### 图计算 graph computing

数据以图为结构的抽象表达，以及在这种数据结构上的计算模式。

### 5 治理与管理

#### 5.1

##### 数据治理 data governance

由一系列旨在设计、实施和监控数据资产管理战略计划的活动来协调和实施的特性或能力。

注：1. ISO/IEC 38505-1 Information technology — Governance of IT — Governance of data 对“数据治理”进行了描述。

2. 数据资产应理解为对组织具有实际利益或潜在利益的一组数据项或数据实体。数据资产是 ISO 55000 中 3.2.1 定义的资产的子集。对组织的好处之一是有利于从分析系统中推导出可付诸行动的知识，因为人们了解到数据具有以前通常未考虑的潜在利益，故通常将其归因于大数据。

3. 数据资产管理战略计划是一个文档，规定如何将数据管理与组织战略保持一致。从数据视角看，该术语与 ISO 55000 Asset management — Overview, principles and terminology 中 3.3.2 定义的战略资产管理计划 (strategic asset management plan, SAMP) 具有相同的含义。

4. GB/T 34960.5—2018《信息技术服务 治理 数据治理规范》中 3.1 对数据治理的定义为“数据资源及其应用过程中相关管控活动、绩效和风险管理的集合”。

[来源：ISO/IEC 20547—3:2020, 3.7, 有修改]

#### 5.2

##### 治理要素 governance element

实施数据治理应关注的关键治理对象或关键治理过程。

[来源：GB/T 34960.1—2017, 3.3, 有修改]

#### 5.3

##### 治理域 governance domain

按照特定层次和功能划分的治理范围，是治理要素的集合。

[来源：GB/T 34960.1—2017, 3.4, 有修改]

#### 5.4

##### 数据管理 data management

通过遵循数据管理评估战略计划，旨在实现最能满足业务目标的大数据架构的一组活动。

注：GB/T 34960.5—2018《信息技术服务 治理 数据治理规范》中 3.2 对数据管理的定义为“数据资源获取、控制、价值提升等活动的集合”。

[来源：ISO/IEC 20547—3:2020, 3.15]

#### 5.5

##### 数据标准管理 data standard management

制定、更新、贯彻、检查和废止数据标准的活动。

5.6

**数据标准 data standard**

数据的命名、定义、结构和取值规范方面的规则和基准。

[来源：GB/T 36344—2018, 2.8]

5.7

**数据质量 data quality**

在特定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

[来源：GB/T 25000.24—2017, 4.11]

5.8

**数据质量管理 data quality management**

管理和控制组织的数据质量的协调活动。

[来源：ISO 8000-2:2020, 3.8.2]

5.9

**质量模型 quality model**

所定义的一组特性及特性间的关系，为规定质量需求和评价质量而提供的一个框架。

[来源：GB/T 25000.24—2017, 4.33]

5.10

**测度 measure**

变量，为其赋值作为测量的结果。

注：术语“测度”可用于对基本测度、派生测度和指标的统称。

[来源：GB/T 25000.24—2017, 4.26]

5.11

**测量 measurement**

以确定测度的值为目的进行的一组操作。

[来源：GB/T 25000.24—2017, 4.27]

5.12

**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、访问权和数据波动性的数据。

[来源：GB/T 5271.17—2010, 17.06.05]

5.13

**元数据管理 metadata management**

制定、更新、贯彻、检查和废止元数据的活动。

5.14

**语义 semantics**

语言的语法成分的含义。

[来源：GB/T 25000.24—2017, 4.35]

5.15

**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。

[来源：GB/T 35295—2017, 2.2.8, 有修改]

## 5.16

**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。

[来源：GB/T 35295—2017，2.1.55]

## 5.17

**数据模型管理 data model management**

收集和了解业务数据需求、制定模型规范、开发数据模型、数据模型应用、符合性检查、模型变更管理等活动的集合。

注：本术语根据GB/T 36073—2018《数据管理能力成熟度评估模型》中8.1.2“过程描述”归纳。

## 5.18

**主数据管理 master data management**

定义主数据的编码规则、数据模型、数据值域、管理流程、质量规则，并实施主数据与应用系统的集成共享等活动的集合。

注：1. 本术语根据 GB/T 36073—2018《数据管理能力成熟度评估模型》中 12.2.2“过程描述”归纳。

2. GB/T 36073—2018《数据管理能力成熟度评估模型》中 12.2.1 对主数据给出的描述是“主数据管理是对主数据标准和内容进行管理，实现主数据跨系统的一致、共享使用的活动”。

## 5.19

**数据安全 data security**

通过采取必要措施，确保数据处于有效保护和合法利用的状态，以及具备保障持续安全状态的能力。

注：GB/T 36073—2018《数据管理能力成熟度评估模型》中3.11对数据安全的定义为“数据的机密性、完整性和可用性”。

## 5.20

**数据安全治理 data security management**

定义数据安全等级、控制数据访问权限，实施用户身份认证和访问行为控制，保护数据安全、管理数据安全风险等活动的集合。

注：1. 本术语根据 GB/T 36073—2018《数据管理能力成熟度评估模型》中 10.2“数据安全治理”归纳。

2. GB/T 36073—2018《数据管理能力成熟度评估模型》中 10.2.1 对数据安全治理给出的概述是“在数据安全标准与策略的指导下，通过对数据访问的授权、分类分级的控制、监控数据的访问等进行数据安全的管理工作，满足数据安全的业务需要和监管需求，实现组织内部对数据生命周期的安全管理”。

## 5.21

**敏感信息 sensitive information**

丢失、滥用、修改或未经授权访问可能导致如下问题的信息：

- a) 对个人的隐私与安全性产生负面影响。
- b) 损害机构的知识产权或商业秘密。
- c) 对国家或机构造成商业损害或经济损失。
- d) 危害国家的安全、内政和外交。

[来源：ISO 19650—5:2020，3.11]

## 5.22

**个人敏感信息 personal sensitive information**

一旦泄露、非法提供或滥用可能危害人身安全和财产安全，极易导致个人名誉、个人身心健康受到损害或受到歧视性待遇等的个人信息。

注：1. 个人敏感信息包括：身份证件号码、个人生物识别信息、银行账户、通信记录和-content、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息、14岁以下（含）儿童的个人信息等。

2. 个人信息控制者通过个人信息或其他信息加工处理后形成的信息，如一旦泄露、非法提供或滥用可能危害人身安全和财产安全，极易导致个人名誉、个人身心健康受到损害或受到歧视性待遇等，属于个人敏感信息。

[来源：GB/T 35273—2020，3.2，有修改]

5.23

**金融业敏感信息** sensitive information in the financial industry

在金融行业中的敏感信息。

5.24

**证券期货业敏感信息** sensitive information in the securities and futures industry

在证券期货业中的敏感信息。

5.25

**匿名化** anonymization

通过对个人信息的技术处理，使得个人信息主体无法被识别或者关联，且处理后的信息不能被复原的过程。

注：个人信息经匿名化处理后的信息不属于个人信息。

[来源：GB/T 35273—2020，3.14]

5.26

**去标识化** de-identification

通过对个人信息的技术处理，使其在不借助额外信息的情况下，无法识别或者关联个人信息主体的过程。

注：去标识化建立在个体基础之上，保留了个体颗粒度，采用假名、加密、哈希函数等技术手段替代对个人信息的标识。

[来源：GB/T 35273—2020，3.15]

5.27

**安全域** security domain

受相同安全策略约束的资产与资源集。

[来源：ISO/IEC 27033-1:2015，3.35]

5.28

**数据生命周期管理** data lifecycle management

对数据生命周期进行管理的活动。

5.29

**数据架构** data architecture

数据要素、结构和接口等抽象及其相互关系的框架。

[来源：GB/T 34960.5—2018，3.5]

5.30

**数据架构管理** data architecture management

对数据架构进行管理的活动。

5.31

**数据资产** data asset

组织拥有和控制的、能够产生效益的数据资源。

[来源：GB/T 34960.5—2018，3.2]

5.32

**数据资产管理** data asset management

组织规划、控制和提供数据资产的一系列业务职能，包括开发、执行和监督有关数据资产的制度、组织及流程中的规划、加工、变更、盘点、引用、处置等活动。

注：数据资产管理的目的是保护和提高数据资产的价值。

## 6 支撑运行环境

### 6.1

#### 架构 **architecture**

在系统及其运行环境中嵌入的元素、关系、设计原则及发展理念类基本概念或属性。

[来源：ISO/IEC/IEEE 42010:2011, 3.2, 有修改]

### 6.2

#### 架构框架 **architecture framework**

在某一应用领域，利益相关方形成的体系结构描述惯例、原则或实践。

示例：1. 广义企业参考架构和方法（Generalised Enterprise Reference Architecture and Methodologies, GERAM）是一种架构框架。

2. 开放分布式处理的参考模型（Reference Model of Open Distributed Processing, RM-ODP）是一种架构框架。

[来源：ISO/IEC/IEEE 42010:2011, 3.4]

### 6.3

#### 架构视图 **architecture view**

从系统特定角度表达系统架构的工作产品。

[来源：ISO/IEC/IEEE 42010:2011, 3.5]

### 6.4

#### 架构观点 **architecture viewpoint**

为构建特定的系统而建立的架构视图惯例，解释和使用约定类工作产品。

[来源：ISO/IEC/IEEE 42010:2011, 3.6]

### 6.5

#### 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) 网络是指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) 计算是指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、逻辑分布等）的软件的物理处理器和储存器。
- c) 储存是指在大数据系统中保存数据的资源（如储存器、本地磁盘、独立磁盘的软硬件冗余阵列、储存域网、附网储存等）。
- d) 环境是指建立大数据系统时要考虑的物理辅助资源（如供电、冷却等）。

2. 是大数据框架提供者可能提供的一种框架。

[来源：GB/T 35295—2017, 2.1.9, 有修改]

### 6.6

#### 数据平台框架 **data platform framework**

用于指导实现结合相关应用编程接口（Application Programming Interface, 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储存或分栏数据储存。

2. 是大数据框架提供者可能提供的一种框架。

[来源：GB/T 35295—2017，2.1.10，有修改]

## 6.7

### 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及定义数据的计算和处理的集合。

注：是大数据框架提供者可能提供的一种框架。

[来源：GB/T 35295—2017，2.1.11，有修改]

## 6.8

### 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应用编程接口的集合。

注：是大数据框架提供者可能提供的一种框架。

[来源：GB/T 35295—2017，2.1.12，有修改]

## 6.9

### 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），从而实现对中央处理器（central processing unit, CPU）和存储两大资源高效且有效管理的集合。

注：是大数据框架提供者可能提供的一种框架。

[来源：GB/T 35295—2017，2.1.13，有修改]

## 6.10

### 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。

注：GB/T 35274—2017《信息安全技术 大数据服务安全能力要求》中3.9对大数据系统的定义为“包括大数据使用者、大数据服务提供者、大数据应用和大数据平台的信息系统”。

[来源：GB/T 35295—2017，2.1.14，有修改]

## 6.11

### 大数据平台 big data platform

采用分布式存储和计算技术，提供大数据的访问和处理，支持大数据应用安全高效运行的软硬件集合，包括监视大数据的存储、输入输出、操作控制等大数据服务软硬件基础设施。

[来源：GB/T 35274—2017，3.6，有修改]

## 6.12

### 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计算节点的文件系统。

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

[来源：GB/T 35295—2017，2.1.21，有修改]

## 6.13

### 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：在大数据卷系统中，在数据读出时启动准备阶段，因此称为“读时模式”。

[来源: GB/T 35295—2017, 2.1.34, 有修改]

#### 6.14

##### 数据库 database

根据概念结构组织的数据合集, 概念结构描述了数据的特性及其相应实体之间的关系, 支持一个或多个应用领域。

[来源: ISO/IEC 20546:2019, 3.1.7]

#### 6.15

##### 关系数据库 relational database

数据按关系模型组织的数据库。

[来源: GB/T 35295—2017, 2.2.5, 有修改]

#### 6.16

##### 非关系数据库 non-relational database

不遵循关系模型组织的数据库。

[来源: ISO/IEC 20546:2019, 3.1.25, 有修改]

#### 6.17

##### 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。

注: 数据仓库是企业所有类型数据有序组织的集合, 为企业制定战略决策提供依据。

[来源: GB/T 35295—2017, 2.1.35, 有修改]

#### 6.18

##### 联合数据库系统 federated database system

一种元数据库管理系统, 可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。

[来源: GB/T 35295—2017, 2.1.39, 有修改]

#### 6.19

##### 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。

示例: 1. 存储域网 (Storage Area Networks, SAN)。

2. 附网存储 (Network Attached Storage, NAS)。

注: 此类系统的技术实现支持同时从多个节点访问多个大型数据集。

[来源: GB/T 35295—2017, 2.1.50, 有修改]

#### 6.20

##### 数据供应链 data supply chain

对大数据服务提供者的数据采集、数据预处理、数据聚合、数据交换、数据访问等相关数据活动进行计划、协调、操作、控制和优化所需的可用数据资源形成的链状结构。

注: 数据供应链的目标是将大数据服务所需的各种数据和系统资产, 通过计划、协调、操作、控制、优化等数据活动, 确保大数据服务提供者能在正确的时间, 按照正确的数据服务协议送给正确的大数据使用者。

[来源: GB/T 35274—2017, 3.10]

#### 6.21

##### 分散—聚集 scatter-gather

大型数据集的处理形式, 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。

示例: MapReduce (包含映射和归约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。

注: 分散—聚集通常要求改变处理软件的算法。

[来源：GB/T 35295—2017，2.1.23，有修改]

6.22

**分布式数据处理 distributed data processing**

在计算机网络中，操作的执行分散在各个节点的数据处理。

注：分布式数据处理需要集体协作，通过节点间的数据通信来实现集体协作。

[来源：ISO/IEC 2382:2015，2120540]

6.23

**集群 cluster**

分布式数据处理中，集中控制下的功能单元集。

[来源：ISO/IEC 2382:2015，2120586]

6.24

**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。

[来源：GB/T 35295—2017，2.1.59]

6.25

**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。

注：1. 水平扩展的示例是通过在集群中添加节点以增加额外资源来提高分布式数据处理的性能。

2. 水平扩展也称作横向扩展或水平伸缩。

[来源：GB/T 35295—2017，2.1.17，有修改]

6.26

**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。

注：垂直扩展也称作垂向扩展或垂直伸缩。

[来源：GB/T 35295—2017，2.1.16，有修改]

6.27

**客户端/服务器架构 client/server architecture**

用于管理共享资源并将这些共享资源的访问作为服务提供给客户端的计算机体系结构。

[来源：ISO 15849:2001，2.6，有修改]

6.28

**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：1. 物即物理实体。

2. 物联网产生的数据是大数据的重要来源。

[来源：GB/T 33745—2017，2.1.1，有修改]

6.29

**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构件失效。

[来源：GB/T 35295—2017，2.1.53，有修改]

附 录  
(资料性)  
本文件中术语描述的策略

本文件的术语按照如下策略进行描述：

- a) 参照国际标准。目前国际标准化组织（ISO）、国际电工委员会（IEC）和国际电信联盟（ITU）等标准化组织均对大数据的标准化予以关注，陆续形成一些标准和技术文件，从不同视角定义了大数据相关概念。本文件充分参考了大数据类国际标准，并直接引入适于金融行业应用的概念。
- b) 赋予金融特征。本文件采用增加限定语、“注”和“示例”的方式对引入的大数据术语赋予了金融特征，便于术语能够在金融领域被统一认知，广泛应用。
- c) 侧重工程实践。本文件注重结合金融行业应用实践对大数据术语进行描述，必要时，增加“注”以进一步说明。
- d) 便于参照查询。本文件力争对大数据某一方面的术语集中描述，或对其某一特性描述尽可能的取值。必要时，对不同来源的术语存有不同释义的情况通过“注”和“示例”以进一步说明。
- e) 易于拓展使用。考虑到实际工作中，常在特定金融语境下使用大数据术语，而金融行业又分为银行、证券、保险等多个子行业，本文件定义的术语力争能够作为专业术语的构词组件，需要时，在术语前面加上必要的限定，即可构成在特定语境下应用的词汇。

行业标准信息平台

## 参 考 文 献

- [1] GB/T 31186—2014 银行客户基本信息描述规范
- [2] GB/T 33745—2017 物联网 术语
- [3] GB/T 35273—2020 信息安全技术 个人信息安全规范
- [4] GB/T 36344—2018 信息技术 数据质量评价指标
- [5] ISO 5127:2017 Information and documentation — Foundation and vocabulary
- [6] ISO 8000:2021 Data quality
- [7] ISO 13008:2012 Information and documentation — Digital records conversion and migration process
- [8] ISO 15704:2019 Enterprise modelling and architecture — Requirements for enterprise-referencing architectures and methodologies
- [9] ISO 15849:2001 Ships and marine technology — Guidelines for implementation of a fleet management system network
- [10] ISO/IEC 15944:2020 Information technology — Business Operational View
- [11] ISO 19650:2020 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling
- [12] ISO 19731:2017 Digital analytics and web analyses for purposes of market, opinion and social research — Vocabulary and service requirements
- [13] ISO 20252:2019 Market, opinion and social research, including insights and data analytics — Vocabulary and service requirements
- [14] ISO 20534:2018 Industrial automation systems and integration — Formal semantic models for the configuration of global production networks
- [15] ISO/TS 21719:2018 Electronic fee collection — Personalization of on-board equipment (OBE)
- [16] ISO/IEC 24765:2017 Systems and software engineering — Vocabulary
- [17] ISO/TR 25100:2012 Intelligent transport systems — Systems architecture — Harmonization of ITS data concepts
- [18] ISO/IEC 27033:2016 Information technology — Security techniques — Network security
- [19] ISO/IEC/IEEE 42010:2011 Systems and software engineering — Architecture description