



# 中华人民共和国国家标准

GB/T 38666—2020

---

## 信息技术 大数据 工业应用参考架构

Information technology—Big data—Industrial application reference architecture

2020-04-28 发布

2020-11-01 实施

国家市场监督管理总局  
国家标准化管理委员会 发布

目 次

前言 ..... III

1 范围 ..... 1

2 规范性引用文件 ..... 1

3 术语和定义 ..... 1

4 缩略语 ..... 2

5 参考架构 ..... 2

6 参考架构构件功能 ..... 3

6.1 系统协调者 ..... 3

6.2 数据提供者 ..... 3

6.3 应用提供者 ..... 5

6.4 大数据计算框架提供者 ..... 6

6.5 数据消费者 ..... 7

6.6 安全和隐私 ..... 7

6.7 管理 ..... 8



## 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:北京数码大方科技股份有限公司、中国电子技术标准化研究院、清华大学、中国人民大学、上海计算机软件技术开发中心、北京软通智城科技有限公司、九次方大数据信息集团有限公司、勤智数码科技股份有限公司、中兴通讯股份有限公司、北京卡达克数据有限公司、四川久远银海软件股份有限公司、湖北省标准化与质量研究院、北京软件和信息服务交易所有限公司。

本标准主要起草人:齐建军、朱道云、卫凤林、张群、王晨、王建民、杜小勇、陈晋川、郑树泉、栗越妍、何昕、李瑛、李正、刘宇峰、黄峥、楚俊生、方祥毅、王英资、吴高、马万钟、于铁强、刘雪。

# 信息技术 大数据 工业应用参考架构

## 1 范围

本标准给出大数据在工业领域的参考架构,规定了各组成部分(构件)的基本功能。  
本标准适用于工业大数据开发、管理和应用。

## 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是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 35295—2017 信息技术 大数据 术语

GB/T 35589—2017 信息技术 大数据 技术参考模型

## 3 术语和定义

GB/T 35295—2017 界定的以及下列术语和定义适用于本文件。

### 3.1

**工业大数据 industrial big data**

大数据理论及技术在工业领域中的应用。

### 3.2

**大数据工业应用参考架构 industrial big data application reference architecture; IBDRA**

对工业大数据内在要求、设计结构和运行进行开放性探讨的高层概念模型。

注:一般包含系统协调者、数据提供者、应用提供者、框架提供者和数据消费者 5 个逻辑功能构件。

### 3.3

**系统协调者 system orchestrator**

IBDRA 中的一种逻辑功能构件,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。

### 3.4

**数据提供者 data provider**

IBDRA 中的一种逻辑功能构件,将新的数据或信息引入大数据系统。

### 3.5

**应用提供者 application provider**

IBDRA 中的一种逻辑功能构件,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需求。

### 3.6

**框架提供者 framework provider**

IBDRA 中的一种逻辑功能构件,即一种计算框架,在此框架中执行应用转换,同时保护数据完整性和隐私。

### 3.7

**数据消费者 data consumer**

IBDRA 中的一种逻辑功能构件,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。

#### 4 缩略语

下列缩略语适用于本文件。

BOM:物料清单(Bill of Material)

CAD:计算机辅助设计(Computer Aided Design)

CAE:计算机辅助工程(Computer Aided Engineering)

CAM:计算机辅助制造(Computer Aided Manufacturing)

CAPP:计算机辅助工艺规划(Computer Aided Process Planning)

CAX:计算机辅助技术/系统(Computer Aided X)

CRM:客户关系管理(Customer Relationship Management)

DNC:分布式数控设备联网(Distributed Numerical Control)

ERP:企业资源计划(Enterprise Resource Planning)

IBDRA:大数据工业应用参考架构(Industrial Big Data Application Reference Architecture)

MES:制造执行系统(Manufacturing Execution System)

NC:数字控制(Numerical Control)

PLM:产品生命周期管理(Product Lifecycle Management)

POSIX:可移植操作系统接口(Portable Operating System Interface)

SCADA:数据采集与监视控制(Supervisory Control and Data Acquisition)

SCM:供应链管理(Supply Chain Management)

#### 5 参考架构

本标准在 GB/T 35589—2017 定义的大数据参考架构的基础上,针对工业领域应用,进一步细化了数据提供者和数据消费者,明确了“工业”领域的应用提供者。本标准定义的 IBDRA 见图 1。

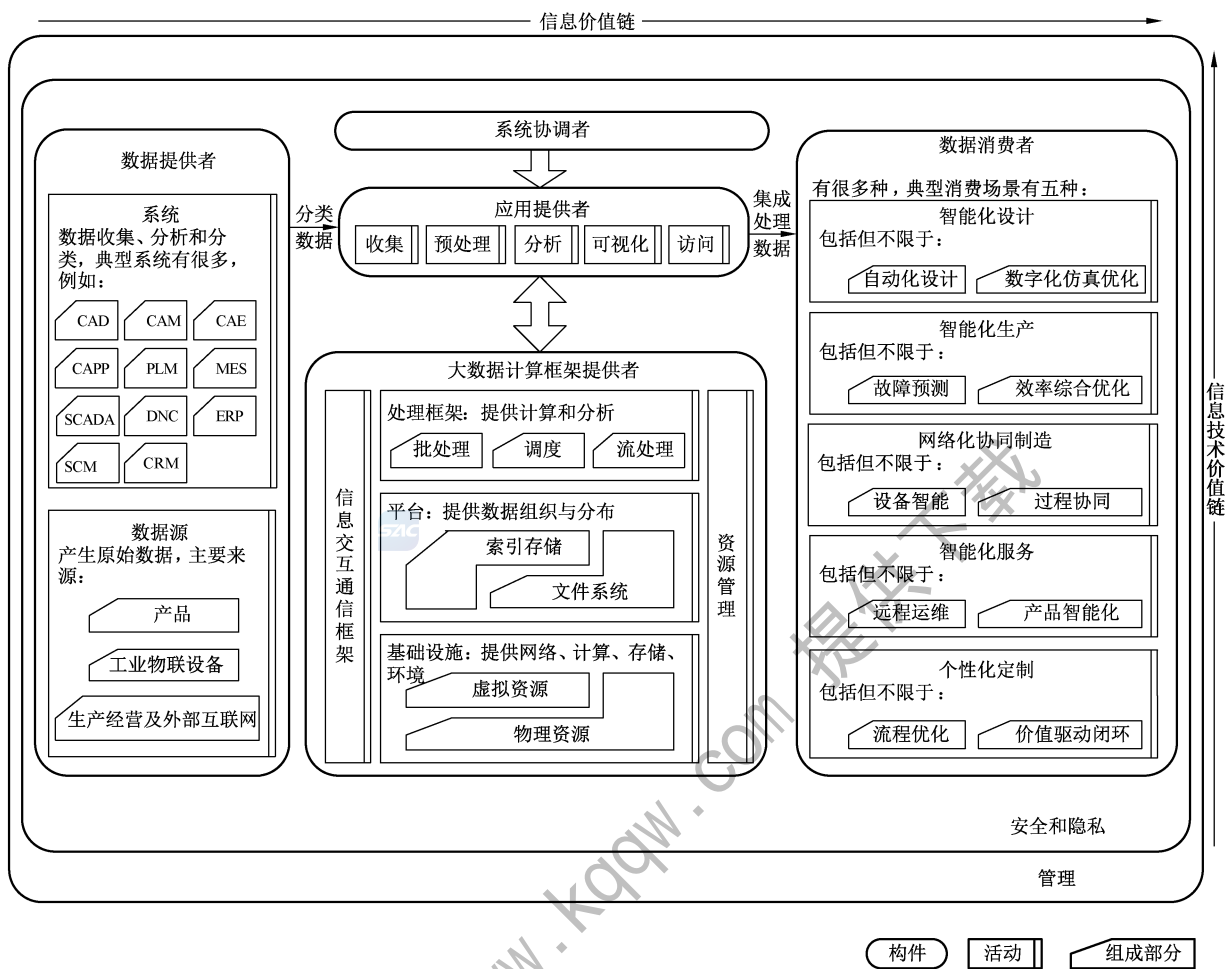


图1 大数据工业应用参考架构

6 参考架构构件功能

6.1 系统协调者

系统协调者的主要功能是规范和集成数据应用活动,具体功能包括四个方面:

- a) 配置和管理 IBDR 中其他构件执行一个或多个工作负载,以确保各工作项能正常运行;
- b) 为其他组件分配对应的物理或虚拟节点;
- c) 对各组件的运行情况进行监控;
- d) 通过动态调配资源等方式来确保各组件的服务质量水平达到所要求。

系统协调者的功能可由管理员、软件或二者的组合以集中式或分布式的形式实现。

6.2 数据提供者

6.2.1 概述

数据提供者的主要功能是将原始数据收集起来,经过预处理提供给工业大数据应用提供者。

此构件主要包括数据源和系统两部分。数据源产生原始数据,之后通过各种信息系统的收集、分析和分类,提供给工业大数据应用提供者。

## 6.2.2 数据源

主要功能是提供原始数据。任何实体以及实体的活动都可能是数据源,例如,各类人员、工业软件、生产设备装备、产品、物联网、互联网、其他软件等各类实体,以及企业活动、人员行为、装备设备运行、环境检测、物联网和互联网运行等各类活动都可能产生数据。图 1 以示例形式列出了以下三大类数据源:

a) 产品:是工业应用数据的核心数据源。

以产品的整个生命周期过程为主线,从时间上覆盖产品市场调研、概念设计、详细设计、工艺设计、生产准备、产品试制、产品定型、产品销售、运行维护、产品报废和回收利用等过程,从空间上覆盖企业内部、供应链上的企业及最终用户,都在产生产品相关数据。这些数据影响许多与产品相关的其他数据源所产生的数据,支持不同的工业应用。

产品相关数据,其表现形式有很多,例如,产品结构和配置、零件定义及设计数据、CAD 三维模型和二维图纸文件、工程分析及验证数据、制造计划及规范、CAD/CAM 编程文件、图像文件(照片、造型图、扫描图等)、产品说明书、软件产品(程序、库、函数等“零部件”)等结构化和非结构数据。具体描述形式取决于设计者设计时的考虑。

b) 工业物联网设备:是工业大数据新的、增长最快的来源。

工业生产设备在物联网运行模式下,产生、收集涵盖操作和运行情况、工况状态、环境参数等数据。

工业物联网设备产生的数据,其表现形式有很多,例如:设备接口、智能传感器产生收集的涵盖设备运行参数、工况状态参数、运行环境参数等评估生产设备运行状态、产品运行状态的数据。具体表现形式取决于设备的功能和使用者的需求。

c) 生产经营及外部互联网:是与工业企业生产活动和产品相关的企业外部互联网,是保障工业应用的支撑数据源。

生产经营及外部互联网数据主要来自传统企业信息化范围,被收集存储在企业信息系统内部。

生产经营及外部互联网数据,其表现形式有很多,例如:企业信息系统累计的产品生产研发数据、试验数据、生产数据、客户信息数据、物流供应数据、环境数据;用户通过互联网发表的与企业产品相关的评论数据、第三方机构提供的用户调查数据、评价企业环境绩效的环境法规、预测产品市场的宏观社会经济数据等。具体表现形式取决于生产经营活动的需求。

## 6.2.3 系统

主要功能是对数据源产生的数据进行收集、分析与分类,然后提供给工业大数据应用提供者。此类系统有很多,主要包括 CAD、CAM、CAE、CAPP、PLM、MES、SCADA、DNC、ERP、SCM、CRM 等。具体如下:

a) CAD 用于生成产品的图形化数据,是所有产品数据的源头,包括非结构化数据(文件)和结构化数据(从文件中提取出的产品属性、BOM 结构和物料清单数据等)。产生的数据可为设计、绘图、工程分析与文档控制等设计活动提供支持。

b) CAM 一般具有数据转换和过程自动化两方面的功能。CAM 在 CAD 模型基础上,生成和仿真优化数控加工用的指令代码数据。产生的 NC 代码可驱动机床设备运行,进行生产设备管理、控制和操作。

c) CAE 一般采用 CAD 系统来建立 CAE 的几何模型和物理模型,完成分析数据的输入,主要处理、分析和优化复杂工程和产品的力学性能等数据。系统产生的结果数据可生成形象的图形输出,为设计活动提供支持。

d) CAPP 通过向计算机输入被加工零件的几何信息(形状、尺寸等)和工艺信息(材料、热处理、批量等),来生成产品和零件的工艺路线、工序内容等加工工艺过程文件。

e) PLM 用于收集和分类管理产品相关结构化和非结构化文件数据,并记录相关角色、环节间的

协同过程数据。

- f) MES 用于收集、管理和优化车间层的制造过程数据。MES 在接收订单开始到制成最终产品的全部时间范围内,采集来自产品、工业物联网设备以及生产经营相关的各种数据信息和状态信息,与上层业务计划层和底层过程控制层进行信息交互。
- g) SCADA 用于收集和管理自动化设备的运行参数、控制、测量及各类信号报警等数据,并向现场连接的设备发送控制命令。
- h) DNC 用于收集、管理和控制数控机床设备的输入、输出数据。其中涉及的数据实体包含四类:
  - 1) 与制造设备硬件相关的数据实体(如机床等);
  - 2) 与人机通信相关的数据实体(如通信协议实体和串口通信实体);
  - 3) 数控数据实体(如 NC 程序号、刀具号、工序号);
  - 4) 输入操作指令或派工单实体。
- i) ERP 围绕企业的业务流程,用于收集和管理企业的物质资源、人力资源、财务资源等数据。
- j) SCM 围绕企业业务协作过程,收集和管理企业与上下游合作伙伴间的协作数据。
- k) CRM 用于收集和管理企业与客户间的交易和服务数据。

### 6.3 应用提供者

#### 6.3.1 概述

应用提供者主要功能是围绕数据消费者需求,将来自数据提供者的数据进行处理和提取,提供给数据消费者。

主要包括收集、预处理、分析、可视化和访问五个活动。

#### 6.3.2 收集

负责处理与数据提供者的接口和数据引入。

由于工业数据的种类、格式很多,且开放程度差异很大,缺少统一标准,需要根据数据格式、类型,通过引用对应的工业应用或构件,完成数据的识别和导入。

#### 6.3.3 预处理

包括数据清洗、数据归约、标准化、格式化和存储。

数据清洗和数据规约是为避免噪声或干扰项给后期分析带来困难,针对首次采集获得的多维异构数据执行的同构化预处理。

数据标准化、格式化处理如:元数据处理和标识管理。具体如下:

- a) 元数据处理包括对订单元数据、产品元数据、供应商能力等进行定义和规范;
- b) 标识管理包括分配与注册、编码分发与测试管理、存储与编码规范、解析机制等。

数据存储主要采用大数据分布式云存储的技术,将预处理后的数据有效存储在性能和容量都能线性扩展的分布式数据库中。

#### 6.3.4 分析

基于数据科学家的需求或垂直应用的需求,利用数据建模、处理数据的算法,以及工业领域专用算法,实现从数据中提取知识的技术。

例如,对无法基于传统建模方法建立生产优化模型的相关工序建立特征模型,基于订单、机器、工艺、计划等生产历史数据、实时数据及相关生产优化仿真数据,采用聚类、分类、规则挖掘等数据挖掘方法及预测机制建立多类基于数据的工业过程优化特征模型。



### 6.3.5 可视化

对经处理、分析运算后的数据,通过合适的显示技术,如大数据可视化技术、工业 2D 或 3D 场景可视化技术等,呈现给最终的数据消费者。

### 6.3.6 访问

与可视化和分析功能交互,响应数据消费者和应用程序的请求。

## 6.4 大数据计算框架提供者

### 6.4.1 概述

按 GB/T 35589—2017 中 7.4.1,大数据框架提供者的主要功能是为工业大数据应用提供者在创建具体应用时提供使用的资源和服务。

大数据计算框架提供者包括基础设施、平台、处理框架、信息通信和资源管理 5 个活动。

### 6.4.2 基础设施

为大数据系统中的所有其他要素提供必要的资源,这些资源是由一些物理资源的组合构成,这些物理资源可以控制/支持相似的虚拟资源。这些资源分为下面四类:

- a) 网络:从一个资源向另一个资源传输数据的资源;
- b) 计算:用于执行和保持其他组件的软件的处理器和存储器;
- c) 存储:大数据系统中保存数据的资源;
- d) 环境:在建立大数据实例的时候需考虑的物理厂房资源(电力、制冷等)。

### 6.4.3 平台

包含逻辑数据的组织和分布,支持文件系统方式存储和索引存储方法:

- a) 文件系统:实施某种级别的 POSIX 标准以获取权限,进行相关的文件操作;
- b) 索引存储:无需扫描整个数据集,便可以迅速定位数据的具体要素。

### 6.4.4 处理框架

提供必要的基础设施软件以支持实现应用程序能够满足数据数量、速度和多样性的处理。包括批处理、流处理,以及两者的数据交换与数据操作。

### 6.4.5 信息通信

包含点对点传输和存储转发两种通信模型。在点对点传输模型中,发送者通过信道直接将所传输的信息发送给接收者;而在后者中,发送者会将信息先发送给中间实体,然后中间实体再逐条转发给接收者。点对点传输模型还包括多播这种特殊的通信模式,在多播中,一个发送者可将信息发送给多个而不是一个接收者。

### 6.4.6 资源管理

计算、存储及实现两者互联互通的网络连接管理。主要目标是实现分布式的、弹性的资源调配,具体包括对存储资源的管理和对计算资源的管理。

## 6.5 数据消费者

### 6.5.1 概述

数据消费者的主要功能是通过调用工业大数据应用提供者提供的接口按需访问信息,并进行加工处理,以达到特定的目标。

数据消费者有很多种,典型的有智能化设计、智能化生产、网络化协同制造、智能化服务和个性化定制五种智能制造模式。

### 6.5.2 智能化设计

以产品数据为核心,通过对输出的产品模型、知识库(例如 2D、3D 图纸,产品结构和工艺路线)、用户使用数据等的集成关联和分析,帮助设计人员实现产品的优化设计、创新设计或自动化设计,典型的智能化设计例如:自动化设计、数字化仿真优化。

自动化设计通过集成工程设计、仿真、试制、试验过程中的多种 CAX 计算机辅助设计工具和系统,实现 CAX 平台数据(例如任务流程数据、工程应用数据、设计知识)的互通;结合智能语义分析,实现设计过程的自动化执行,并在此基础上实现多学科综合设计优化。

数字化仿真优化用于顺应设计数据的关联性,在设计阶段有效地对产品进行综合的评估和改进。

### 6.5.3 智能化生产

指人机智能交互、工业机器人、制造工艺的仿真优化、数字化控制、状态监测等先进工业技术在生产制造中的应用。主要包括生产效率综合优化、生产故障预测等典型场景。

生产效率综合优化通过对生产过程中相关产线、装备、设备的关键指标进行监控、数据挖掘和分析,实现产线升级、产品质量优化、设备故障诊断与维护、智能排程、智能生产等,综合优化生产过程和生产效率。

生产故障预测根据经验数据进行对比建模,对生产过程中的故障进行预测诊断,降低设备故障对制造过程造成的影响。

### 6.5.4 网络化协同制造

在设备物联、智能控制、生产过程透明化基础上,实现制造资源、制造能力、制造过程的信息透明,根据订单特征,连通不同物理区域的多样化生产资源,完成最优化的资源配置,实现高质高效的协同制造、组装和交付过程。

### 6.5.5 智能化服务

在生产管理服务和产品售后服务环节,对产品运行及使用数据进行收集、分析和优化,所得数据结果可以辅助产品设计的优化,以及产品的智能化诊断、运维和远程控制等。同时,可以实现智能检测监管的应用,如危险化学品、食品、印染、稀土、农药等重点行业智能检测监管应用。

### 6.5.6 个性化定制

基于用户定制化需求,通过全流程建模和数据集成贯通,将用户需求和企业产品设计、生产计划精准匹配,并借助模块化产线、新型制造工艺,积累并闭环优化各环节经验数据,实现数据流动的自动化、智能化,并通过数据驱动计划、设计、生产、物流和交付等过程,持续优化提升整体运作效率和用户体验。

## 6.6 安全和隐私

在安全和隐私管理模块,通过不同的技术手段和安全措施,构建大数据平台安全防护体系,实现覆

盖硬件、软件 and 上层应用的安全保护,从网络安全、主机安全、应用安全、数据安全四个方面来保证大数据平台的安全性。

按 GB/T 35589—2017 中 7.6,该构件主要包括以下四种功能:

- a) 网络安全:通过网络安全技术,保证数据处理、存储安全和维护正常运行;
- b) 主机安全:通过对集群内节点的操作系统安全加固等手段保证节点正常运行;
- c) 应用安全:具有身份鉴别和认证、用户和权限管理、数据库加固、用户口令管理、审计控制等安全措施,实施合法用户合理访问资源的安全策略;
- d) 数据安全:从集群容灾、备份、数据完整性、数据分角色存储、数据访问控制等方面保证用户数据的安全。

同时应提供一个合理的灾备框架,提升灾备恢复能力,实现数据的实时异地容灾功能,跨数据中心数据备份。

隐私保护主要是在不暴露用户敏感信息的前提下进行有效的数据挖掘;根据需保护的内容不同,可分为位置隐私保护、标识符匿名保护和连接关系匿名保护等。



## 6.7 管理

按 GB/T 35589—2017 中 7.7,该构件的主要功能覆盖以下各方面:

- a) 提供大规模集群统一的运维管理系统,能够对包括数据中心、基础硬件、平台软件和应用软件进行集中运维、统一管理,实现安装部署、参数配置、监控、告警、用户管理、权限管理、审计、服务管理、健康检查、问题定位、升级和修复等功能;
- b) 具有自动化运维的能力,通过对多个数据中心的资源进行统一管理,合理地分配和调度业务所需要的资源,做到自动化按需分配。同时提供对多个数据中心的信息技术基础设施进行集中运维的能力,自动化监控数据中心内各种信息技术设备的事件、告警、性能,实现从业务纬度来进行运维的能力;
- c) 对主管理系统节点及所有业务组件中心管理节点实现高可靠性的双机机制,采用主备或负荷分担配置,避免单点故障场景对系统可靠性的影响。