

# 工业大数据白皮书

## (2019 版)

主编单位： 中国电子技术标准化研究院  
全国信息技术标准化技术委员会大数据标准工作组

联合主编单位： 工业大数据产业应用联盟

二零一九年三月

## 专家指导组：

梅 宏 林 宁 杜小勇 吴建明 闵万里 田日辉

## 主 编：

孙文龙

## 副主编：

代 红 吴东亚 王 晨 程宏斌 张 巍 于辰涛

## 编写人员（排名不分先后）

尹 卓 张 群 陈宏志 蔡春久 于 洋 魏 清 黄先芝 闵京华  
卫凤林 齐建军 朱道云 徐 哲 孙 川 李三华 禹晓庆 倪 岭  
赵志强 王慧生 李千目 侯 君 戚 湧 包建国 梅 军 鲁 梨  
王 飞 冀占峰 辛颖梅 仇卫文 张海静 袁海飞 付思敏 姜 波  
吴小伟 唐怀坤 田岩岩 吕劲松 鲍广宇 倪晓炜 刘 婵 刘振宇  
卞志刚 胡杰英 刘清华 王计斌 郭一新

## 编写单位（排名不分先后）

中国电子技术标准化研究院	北京工业大数据创新中心
清华大学	徐工集团工程机械股份有限公司
美林数据技术股份有限公司	浪潮软件集团有限公司
联想（北京）有限公司	江苏蓝创智能科技股份有限公司
江苏省工业和信息化厅	中电长城网际系统应用有限公司
石化盈科信息技术有限责任公司	北京东方国信科技股份有限公司
江苏中堃数据技术有限公司	南京理工大学
南京天数智芯科技有限公司	北京数码大方科技股份有限公司
南京擎天科技有限公司	上海智能制造系统创新中心有限公司
南京华苏科技有限公司	朗坤智慧科技股份有限公司
中通服咨询设计研究院有限公司	江苏思特瑞信息技术有限公司
深圳迅策科技有限公司	上海智能制造功能平台有限公司
航天云网科技发展有限责任公司	南京维拓科技股份有限公司
上海计算机软件发展中心	杉数（北京）科技有限公司
网易（杭州）网络有限公司	徐州徐工挖掘机械有限公司
江苏蓝雅信息科技有限公司	

## 版权声明

本白皮书版权属于中国电子技术标准化研究院，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，请注明“来源：中国电子技术标准化研究院”。违反上述声明者，本院将追究其相关法律责任。

中国电子技术标准化研究院

# 目 录

1 前言.....	1
1.1 研究背景.....	1
1.2 研究目标及意义.....	1
2 工业大数据内涵.....	3
2.1 工业大数据的定义与边界.....	3
2.1.1 工业大数据的定义.....	3
2.1.2 工业大数据的边界.....	3
2.2 与智能制造的关系.....	4
2.2.1 在智能制造中的应用.....	4
2.2.2 在智能制造标准体系中的定位.....	5
2.3 与工业互联网的关系.....	7
2.3.1 在工业互联网中的应用.....	8
2.3.2 在工业互联网标准体系中的定位.....	9
3 工业大数据产业发展现状.....	10
3.1 政策支撑.....	10
3.1.1 国际工业大数据战略.....	10
3.1.2 国内工业大数据政策.....	11
3.2 工业大数据应用参考架构.....	14
3.3 典型应用场景.....	17
3.3.1 智能化设计.....	18

3.3.2	智能化生产	19
3.3.3	网络化协同制造	21
3.3.4	智能化服务	22
3.3.5	个性化定制	23
3.4	产业发展分析	24
4	工业大数据关键技术	25
4.1	工业大数据技术架构	25
4.2	工业大数据平台	27
4.3	工业大数据采集技术	28
4.4	工业大数据存储与管理技术	29
4.4.1	多源异构数据管理技术	29
4.4.2	多模态数据集成技术	31
4.5	工业大数据分析技术	32
4.5.1	时序模式分析技术	33
4.5.2	工业知识图谱技术	34
4.5.3	多源数据融合分析技术	34
4.6	工业大数据的前沿技术趋势	35
5	工业大数据管理	36
5.1	数据管理体系	36
5.2	数据资源编目	37
5.2.1	数据资源目录应用架构	37
5.2.2	数据资源编目	38

5.3 数据质量管理.....	39
5.3.1 工业大数据质量问题.....	39
5.3.2 工业大数据质量检测方法.....	41
5.3.3 工业大数据质量管理.....	41
5.4 主数据管理.....	42
5.4.1 主数据管控体系.....	43
5.4.2 主数据应用管理.....	45
5.5 数据安全.....	46
5.5.1 工业大数据安全问题.....	46
5.5.2 工业大数据安全防护技术.....	46
5.5.3 工业大数据安全管理机制.....	48
6 工业大数据的标准体系.....	50
6.1 工业大数据标准化工作基础.....	50
6.1.1 国际标准化.....	50
6.1.2 国内标准化.....	50
6.2 工业大数据标准体系框架.....	52
6.3 工业大数据标准明细表.....	54
6.4 工业大数据重点标准描述.....	55
7 下一步工作建议.....	56

# 1 前言

## 1.1 研究背景

当前，全球主要国家掀起了新一轮以“信息技术与制造业融合”为共同特征的工业革命，加速发展新一代信息技术，并推动其与全球工业系统的深度融合，以期抢占新一轮产业竞争的制高点。无论是欧美老牌国家制造业的重振，还是中国制造业的转型提升，工业大数据都将发挥不可替代的作用。习近平总书记在党的十九大报告中提出“加快建设制造强国，加快发展先进制造业，推动互联网、大数据、人工智能和实体经济深度融合，在中高端消费、创新引领、绿色低碳、共享经济、现代供应链、人力资本服务等领域培育新增长点、形成新动能。”工业大数据的创新发展，必将成为未来提升制造业生产力和竞争力的关键要素，是目前全球工业转型面临的重大课题。

全国信息技术标准化技术委员会大数据标准工作组由工信部和国家市场监督管理总局共同指导，主要承担我国大数据领域标准化工作，已经发布《大数据标准化白皮书》、《工业大数据白皮书（2017版）》等研究成果。为充分发挥国家标准的规划与指导作用，促进我国工业大数据标准体系建设，推动工业大数据在智能制造、工业互联网中的发展应用，全国信标委大数据标准工作组联合工业大数据产业应用联盟启动《工业大数据白皮书（2019版）》编制工作，在2017版的基础上更新了工业大数据发展的最新趋势以及在重点领域的应用实践，提出工业大数据管理的相关方法，进一步完善了工业大数据标准体系，给出工业大数据标准化工作建议，我国工业大数据相关标准研究成果在本版本中也有所体现。

## 1.2 研究目标及意义

本白皮书的目标在于梳理工业大数据技术、产业发展现状，通过综述全球主要国家在工业大数据上的战略部署，从典型应用、技术架构的角度进行分析，勾画出工业大数据发展的整体轮廓，并结合国内工业大数据的发展态势，总结我国在推进工业大数据发展过程中存在的问题及难点，为下一步工作理清思路。



此外，本白皮书围绕“工业数据管理”这一重要议题，提炼了当前业界在工业数据管理体系、工业数据资源编目、工业数据质量管理、工业数据主数据管理、工业数据安全等方面的管理方法。最后基于工业大数据技术、产业发展需求，完善了工业大数据标准体系，为推动工业大数据落地应用和战略部署提供标准化支撑。

中国电子技术标准化研究院

## 2 工业大数据内涵

### 2.1 工业大数据的定义与边界

#### 2.1.1 工业大数据的定义

工业大数据是指在工业领域中，围绕典型智能制造模式，从客户需求到销售、订单、计划、研发、设计、工艺、制造、采购、供应、库存、发货和交付、售后服务、运维、报废或回收再制造等整个产品全生命周期各个环节所产生的各类数据及相关技术和应用的总称。工业大数据以产品数据为核心，极大延展了传统工业数据范围，同时还包括工业大数据相关技术和应用。

工业大数据具备双重属性：价值属性和产权属性。一方面，通过工业大数据分析等关键技术能够实现设计、工艺、生产、管理、服务等各个环节智能化水平的提升，满足用户定制化需求，提高生产效率并降低生产成本，为企业创造可量化的价值；另一方面，这些数据具有明确的权属关系和资产价值，企业能够决定数据的具体使用方式和边界，数据产权属性明显。工业大数据的价值属性实质上是基于工业大数据采集、存储、分析等关键技术，对工业生产、运维、服务过程中数据实现价值的提升或变现；工业大数据的产权属性则偏重于通过管理机制和管理方法帮助工业企业明晰数据资产目录与数据资源分布，确定所有权边界，为其价值的深入挖掘提供支撑。本白皮书在第 4、5 章节分别对工业大数据关键技术和工业大数据管理体系进行介绍。

#### 2.1.2 工业大数据的边界

工业大数据的边界可以从数据来源、工业大数据的应用场景两大维度进行明确。从数据的来源看，工业大数据主要包括三类：

第一类是企业运营管理相关的业务数据。这类数据来自企业信息化范畴，包括企业资源计划(ERP)、产品生命周期管理(PLM)、供应链管理(SCM)、客户关系管理(CRM)和能耗管理系统(EMS)等，此类数据是工业企业传统意义上的数据资

产。

第二类是制造过程数据。主要是指工业生产过程中，装备、物料及产品加工过程的工况状态参数、环境参数等生产情况数据，通过 MES 系统实时传递，目前在智能装备大量应用的情况下，此类数据量增长最快。

第三类是企业外部数据。包括工业企业产品售出之后的使用、运营情况的数据，同时还包括大量客户名单、供应商名单、外部的互联网等数据。

从工业大数据的应用场景看，工业大数据是针对每一个特定工业场景，以工业场景相关的大数据集为基础，集成工业大数据系列技术与方法，获得有价值信息的过程。工业大数据应用的目标是从复杂的数据集中发现新的模式与知识，挖掘得到有价值的信息，从而促进工业企业的产品创新、运营提质和管理增效。根据行业自身的生产特点和发展需求，工业大数据在不同行业中的应用重点以及所产生的业务价值也不尽相同。在流程制造业中，企业利用生产相关数据进行设备预测性维护、能源平衡预测及工艺参数寻优，可以降低生产成本、提升工艺水平、保障生产安全。对于离散制造业，工业大数据的应用促进了智慧供应链管理、个性化定制等新型商业模式的快速发展，有助于企业提高精益生产水平、供应链效率和客户满意度。

## 2.2 与智能制造的关系

智能制造是工业大数据的载体和产生来源，其各环节信息化、自动化系统所产生的数据构成了工业大数据的主体。另一方面，智能制造又是工业大数据形成的数据产品最终的应用场景和目标。工业大数据描述了智能制造各生产阶段的真实情况，为人类读懂、分析和优化制造提供了宝贵的数据资源，是实现智能制造的智能来源。工业大数据、人工智能模型和机理模型的结合，可有效提升数据的利用价值，是实现更高阶的智能制造的关键技术之一。

### 2.2.1 在智能制造中的应用

狭义的智能制造（Smart Factory）主要针对制造业企业的生产过程，从工业 2.0、工业 3.0 到工业 4.0 的进阶过程中，首先是关注提升系统的自动化水平，

完善 MES、APS 等信息化系统的建设，面向对整个生产过程的流程优化实现提质增效。同时，整个生产体系的数字化水平得到极大提升，使得从生产设备、自动化系统、信息化系统中提取数据对人、机、料、法、环等生产过程关键要素进行定量刻画、分析成为可能。这既是从自动化、信息化走向智能化目标的过程，也是通过数字化、网络化最终实现智能化的现实路径。

智能化（Intelligent）描述了自动化与信息化之上的智能制造的愿景，通过对工业大数据的展现、分析和利用，可以更好地优化现有的生产体系：通过对产品生产过程工艺数据和质量数据的关联分析，实现控制与工艺调整优化建议，从而提升产品良率；通过零配件仓储库存、订单计划与生产过程数据分析，实现更优的生产计划排程；通过对生产设备运行及使用数据的采集、分析和优化，实现设备远程点检及智能化告警、智能健康检测；通过对耗能数据的监测、比对与分析，找到管理节能漏洞、优化生产计划，实现能源的高效使用等。

更为广义的智能制造本质是数据驱动的创新生产模式，在产品市场需求获取、产品研发、生产制造、设备运行、市场服务直至报废回收的产品全生命周期过程中，甚至在产品本身的智能化方面，工业大数据都将发挥巨大的作用。例如，在产品的研发过程中，将产品的设计数据、仿真数据、实验数据进行整理，通过与产品使用过程中的各种实际工况数据的对比分析，可以有效提升仿真过程的准确性，减少产品的实验数量，缩短产品的研发周期。再如，在产品销售过程中，从源头的供应商服务、原材料供给，到排产协同制造，再到销售渠道和客户管理，工业大数据在供应链优化、渠道跟踪和规划、客户智能管理等各方面，均可以发挥全局优化的作用。在产品本身的智能化方面，通过产品本身传感数据、环境数据的采集、分析，可以更好的感知产品所处的复杂环境与工况，以提升产品效能、节省能耗、延长部件寿命等优化目标为导向，在保障安全性的前提下，实现在边缘侧对既定的控制策略提出优化建议或者直接进行一定范围内的调整。

## 2.2.2 在智能制造标准体系中的定位

智能制造标准体系如图 2-1 所示。工业大数据标准属于智能制造标准体系“智能赋能技术”部分，为智能制造提供技术和数据支撑。

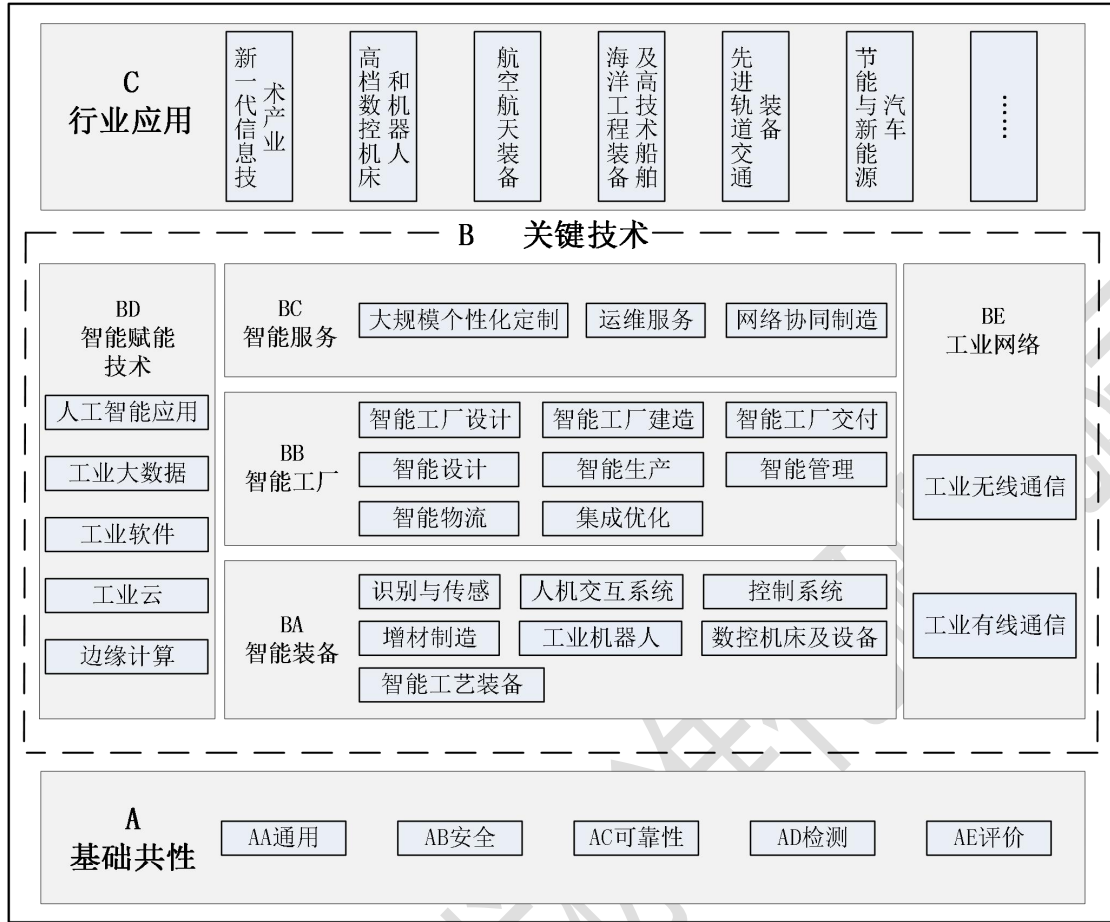


图 2-1 工业大数据标准在智能制造标准化体系中的定位

工业大数据标准在智能制造标准体系框架中的具体位置如图 2-2 所示。

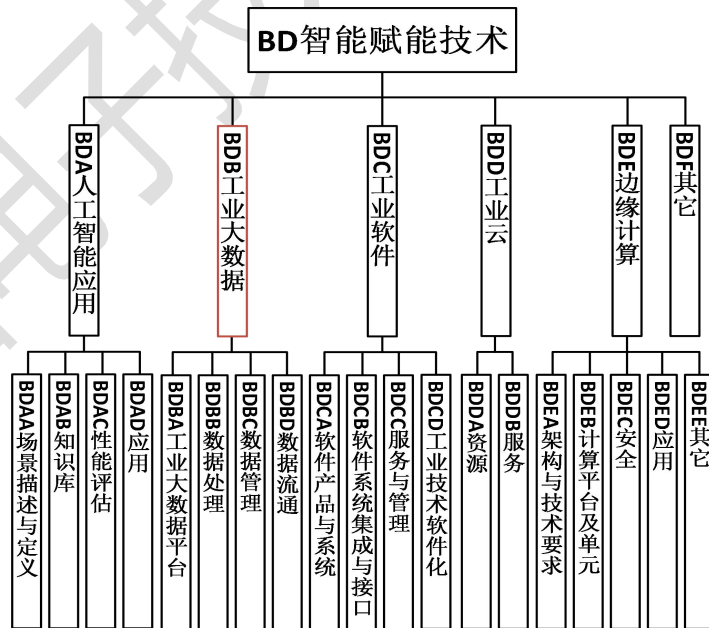


图 2-2 智能制造标准体系-智能赋能技术标准

《国家智能制造标准体系建设指南（2018年版）》中关于智能制造标准体

系中的工业大数据标准给出了具体的描述：主要包括平台建设的要求、运维和检测评估等工业大数据平台标准；工业大数据采集、预处理、分析、可视化和访问等数据处理标准；数据质量、数据管理能力等数据管理标准；工厂内部数据共享、工厂外部数据交换等数据流通标准。主要用于典型智能制造模式中，提高产品全生命周期各个环节所产生的各类数据的处理和应用水平。

## 2.3 与工业互联网的关系

与智能制造的场景有所区别，工业互联网更为关注制造业企业如何以工业为本，通过“智能+”打通、整合、协同产业链，催生个性化定制、网络化协同、服务化延伸等新模式，从而提升企业、整体行业价值链或是区域产业集群的效率。与智能制造相似的，工业互联网既是工业大数据的重要来源，也是工业大数据重要的应用场景。尤其在工业互联网平台的建设中，工业大数据扮演着重要的角色。

2017年，工业和信息化部指导编写了《工业互联网平台白皮书》，围绕工业互联网平台的概念、技术体系、产业生态、应用场景等进行了研究，提出了工业互联网平台功能架构图，如图2-3所示。



图 2-3 工业互联网平台功能架构图

在工业互联网平台功能架构中，工业大数据技术、工业大数据系统是工业互

联网平台层（工业 PaaS 层）的重要核心。一方面，借助工业大数据处理、预处理、分析等技术，基于工业大数据系统，平台层（工业 PaaS 层）得以实现对边缘层、IaaS 层产生的海量数据进行高质量存储与管理；另一方面通过工业大数据建模、分析、可视化等技术，将数据与工业生产实践经验相结合，构建机理模型，支撑应用层各种分析应用的实现。

### 2.3.1 在工业互联网中的应用

工业大数据在工业互联网中的应用首先体现在对于工业互联网个性化定制、网络化协同、服务化延伸等工业互联网新模式场景的支撑。在大规模个性化定制场景下，企业通过外部平台采集客户个性化需求数据，与工业企业生产数据、外部环境数据相融合，建立个性化产品模型，将产品方案、物料清单、工艺方案通过制造执行系统快速传递给生产现场，进行生产线调整和物料准备，快速生产出符合个性化需求的定制化产品。在网络化协同场景下，基于工业大数据，驱动制造全生命周期从设计、制造到交付、服务、回收各个环节的智能化升级，最终推动制造全产业链智能协同，优化生产要素配置和资源利用，消除低效中间环节，整体提升制造业发展水平和世界竞争力。在服务化延伸场景中，通过传感器和工业大数据分析技术，对产品使用过程中的自身工作状况、周边环境、用户操作行为等数据进行实时采集、建模、分析，从而实现在线健康检测、故障诊断预警等服务，催生支持在线租用、按使用付费等新的服务模型，创造产品新的价值，实现制造企业的服务化转型。

除了在工业互联网新模式场景中的应用，从集中化平台的角度来看，工业互联网平台还承载了通过工业大数据的分析利用从而实现知识积累的重任。工业领域经历了数百年的发展，在不同的行业、领域和场景下积累了大量的工业机理和工业知识，体现了对工业过程的深刻理解，能够持续地指导工业过程的优化和改进。在工业大数据时代，通过对这些工业机理、知识的提炼和封装，实现工业机理、知识模型上云、共享和复用，一方面，将使工业机理更好地融入于工业大数据算法，实现模型的调优和迭代，缩短数据模型的收敛时间；同时，通过对海量工业大数据的深入挖掘、提炼、建模和封装，进一步形成面向各个细分工业领域的各类知识库、工具库、模型库和工业软件，将有助于加速旧知识的复用和新知

识的不断产生，进一步服务于工业过程的改进和提升，为用户提供基于工业互联网的持续价值创造良性闭环。

### 2.3.2 在工业互联网标准体系中的定位

2019年，工业和信息化部、国家标准化委员会共同组织制定并发布了《工业互联网综合标准化体系建设指南》。文中提出了工业互联网标准体系框架，包括基础共性、总体、应用三大类标准，工业大数据标准在工业互联网标准体系中处于“平台与数据”部分，如图2-4所示。

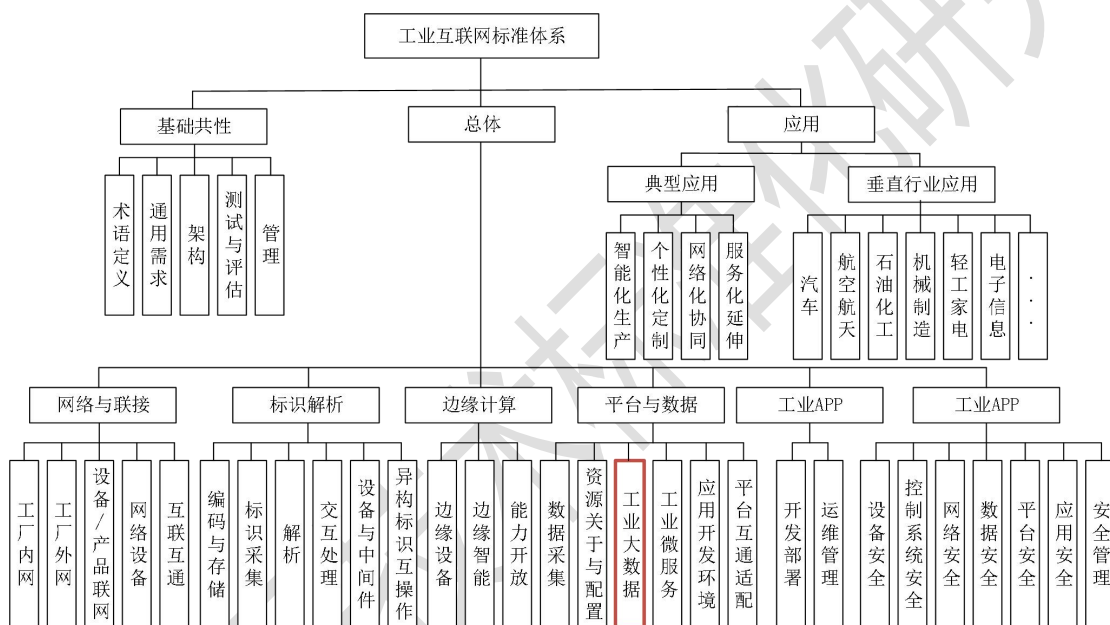


图 2-4 工业互联网标准体系框架

在工业互联网标准体系框架中，工业大数据标准主要包括工业数据交换标准，用于规范工业互联网平台内不同系统之间数据交换体系架构、互操作、性能等要求；工业数据分析与系统标准，用于规范工业互联网数据分析的流程及方法；工业数据管理标准，用于规范工业互联网数据的存储结构、数据字典、元数据、数据质量要求、数据生命周期管理、数据管理能力成熟度等要求；工业数据建模标准，用于规范物理实体（在制品、设备、产线、产品等）在网络空间中的映像及相互关系；以及工业大数据服务标准，用于规范工业互联网平台运用大数据能力对外提供的服务，包括大数据存储服务、大数据分析服务、大数据可视化服务、数据建模及数据开放、数据共享等相关标准。



## 3 工业大数据产业发展现状

### 3.1 政策支撑

#### 3.1.1 国际工业大数据战略

工业是国民经济的基础和支柱，也是一国经济实力和竞争力的重要标志。随着云计算、大数据和物联网等新兴技术的发展，全球掀起了以制造业转型升级为首要任务的新一轮工业变革，主要的工业发达体纷纷制定工业再发展战略。

##### (1) 美国

2014年，美国白宫总统行政办公室发布《2014年全球大数据白皮书》，文中指出，美国大型企业在投资大数据科技方面存在以下几个关键驱动因素：分析运营和交易数据的能力；洞察客户线上消费的行为，以向市场提供新的高度复杂的产品；对组织中的机器和设备进行更加深入的感知。

2018年10月，美国白宫发布了四年一度的《美国先进制造领导战略》，在“智能数字制造”部分提出下一步计划“要通过将大数据分析和先进的传感和控制技术应用于大量制造业活动，促进制造业的数字化转型”。

##### (2) 德国

2015年4月，德国提出“工业4.0”战略。强调通过信息网络与工业生产系统的充分融合，使产品与生产设备之间、工厂内部纵向之间、工厂与工厂之间，都能通过CPS（信息物理系统）联结为一个整体，从而实现生产的智能化，提升制造业的灵活性和工程效率。工业4.0战略展现了一幅全新的工业蓝图：在现实和虚拟结合的网络世界里，互联网将渗透到所有的关键领域，价值创造过程将会改变，原有的行业界限将会消失，新兴的产业链条将会重组，全新的商业模式和合作模式将会出现。

德国“工业4.0”战略的实施重点在于信息互联技术与传统工业制造的结合，其中大数据分析作为关键技术将得到较大范围应用。一是“智能工厂”，重点研

究智能化生产系统及过程,以及网络化分布式生产设施的实现;二是“智能生产”,主要涉及整个企业的生产物流管理、人机互动以及 3D 技术在工业生产过程中的应用等;三是“智能物流”,主要通过互联网、物联网、物流网,整合物流资源,充分发挥现有物流资源供应方的效率,需求方则能够快速获得服务匹配,得到物流支持。

### (3) 法国

2015 年,法国推出“新工业法国战略”,总体布局为“一个核心,九大支点”。一个核心即“未来工业”,主要内容是实现工业生产向数字化、智能化转型,以生产工具的转型升级带动商业模式转型。九大支点,包括新资源开发、可持续发展城市、环保汽车、网络技术、大数据技术、新型医药等,一方面旨在为“未来工业”提供支撑,另一方面重在满足人们日常生活的新需求。该战略为期十年,主要解决三大问题:能源、数字革命和经济生活。

2015 年 5 月,法国经济、工业与就业部又公布了未来工业计划,该计划将在“新工业战略”的第二阶段中扮演核心角色,主要目标是建立更为互联互通、更具有竞争力的法国工业,旨在使工业工具更加现代化,并通过数字技术帮助企业转变经营模式、组织模式、研发模式和商业模式,实现经济增长模式转变。未来工业计划提倡在一些优先领域发展工业模式,例如新资源、可持续发展城市、未来交通、未来医药、数字经济、智能物体、数字安全和智能电网等。

### 3.1.2 国内工业大数据政策

近年来,工业大数据作为我国“智能制造”和“工业互联网”的关键技术支撑以及两化融合的重要基础备受关注。党中央、国务院出台了一系列“大数据”、“两化融合”、“互联网与制造业融合”等综合性政策与指示,其中对工业大数据发展提出了明确的要求,全面指导我国工业大数据技术发展、产业应用及其标准化进程。

表 3-1 国家工业大数据相关政策

排序	政策名称	发布日期	发文单位
1	国务院关于积极推进“互联网+”行动的指导意见	2015年07月	国务院
2	国务院关于印发促进大数据发展行动纲要的通知	2015年08月	国务院
3	国务院关于深化制造业与互联网融合发展的指导意见	2016年05月	国务院
4	关于深化“互联网+先进制造业”发展工业互联网的指导意见	2017年11月	国务院

2015年7月，国务院下发《关于积极推进“互联网+”行动的指导意见》，进一步推动互联网与制造业融合，提升制造业数字化、网络化、智能化水平。文中指出要加强产业链协作，在重点领域推进智能制造、大规模个性化定制、网络化协同制造和服务型制造，打造一批网络化协同制造公共服务平台，加强工业大数据的开发与利用，支撑制造业智能化转型，构建开放、共享、协作的智能制造产业生态。

2015年8月，国务院印发《促进大数据发展行动纲要》，在“工业和新兴产业大数据工程”专栏中指出发展工业大数据，推动大数据在工业研发设计、生产制造、经营管理、市场营销、售后服务等产品全生命周期、产业链全流程各环节的应用，分析感知用户需求，提升产品附加价值。同时提到要建立面向不同行业、不同环节的工业大数据资源聚合和分析应用平台，抓住互联网跨界融合机遇，促进大数据、物联网、云计算和三维（3D）打印技术、个性化定制等在制造业全产业链集成运用，推动制造模式变革和工业转型升级。

2016年5月，《国务院关于深化制造业与互联网融合发展的指导意见》中提出要推动制造企业与互联网企业在发展理念、产业体系、生产模式、业务模式等方面全面融合，深化工业云、大数据等技术的集成应用，汇聚众智，加快构建新型研发、生产、管理和服务模式，促进技术产品创新和经营管理优化，提升企业整体创新能力和水平。

2017年11月27日，国务院《关于深化“互联网+先进制造业”发展工业互联网的指导意见》中再次指出要加快建设和发展工业互联网，推动互联网、大数据、人工智能和实体经济深度融合，发展先进制造业，支持传统产业优化升级，工业大数据在制造业的应用进一步贯彻落实。

围绕国家政策我国工业和信息化部也出台了一系列文件用以促进和推动工业大数据在工业领域中的快速应用。

表 3-2 工业和信息化部相关政策

排序	政策名称	发布日期	发文单位
1	信息化和工业化融合发展规划(2016-2020年)的通知	2015年5月	工业和信息化部
2	大数据产业规划(2016-2020年)	2017年1月	工业和信息化部
3	工业互联网发展行动计划(2018-2020年)	2018年6月	工业和信息化部

2015年5月,工业和信息化部在《信息化和工业化融合发展规划(2016-2020年)》中指出,要推广网络化生产新模式,引领生产方式持续变革;同时深化物联网标识解析、工业云服务、工业大数据分析等在重点行业应用,构建智能监测监管体系,支持机械、汽车等行业发展产品在线维护、远程运维、智能供应链、协同研发等服务新业态,推动大数据在工业设计、生产制造、售后服务等产品全生命周期的应用。

2017年1月,工业和信息化部在《大数据产业规划(2016-2020年)》中指出要以应用为导向,突破大数据关键技术,推动产品和解决方案研发及产业化,创新技术服务模式,形成技术先进、生态完备的技术产品体系。加强工业大数据基础设施建设规划与布局,推动大数据在产品全生命周期和全产业链的应用,推进工业大数据与自动控制和感知硬件、工业核心软件、工业互联网、工业云和智能服务平台融合发展,形成数据驱动的工业发展新模式,探索建立工业大数据中心。

2018年6月,工业和信息化部发布《工业互联网发展行动计划(2018-2020年)》,文中在重点任务“工业互联网平台建设行动”中明确提到“支持建设涵盖基础及创新技术服务、监测分析服务、工业大数据管理、标准管理服务等的平台公共支撑体系”。

此外,工信部连年来通过遴选工业大数据相关产业发展试点示范项目,以及“智能制造综合标准化项目”《工业大数据服务标准化和试验验证系统》等创新应用新模式项目申报及建设支持,围绕深化制造业与互联网融合发展,以及大数据技术在制造业中的深度应用,不断刺激产业进步,增强制造业转型升级新动能,已促成一批高质量行业应用级工业大数据平台的落地,助推工业企业转型发展。

## 3.2 工业大数据应用参考架构

结合《大数据标准化白皮书（2018版）》中提出的大数据参考架构，针对工业领域的应用，本白皮书给出了工业大数据应用参考架构如图 3-1 所示。

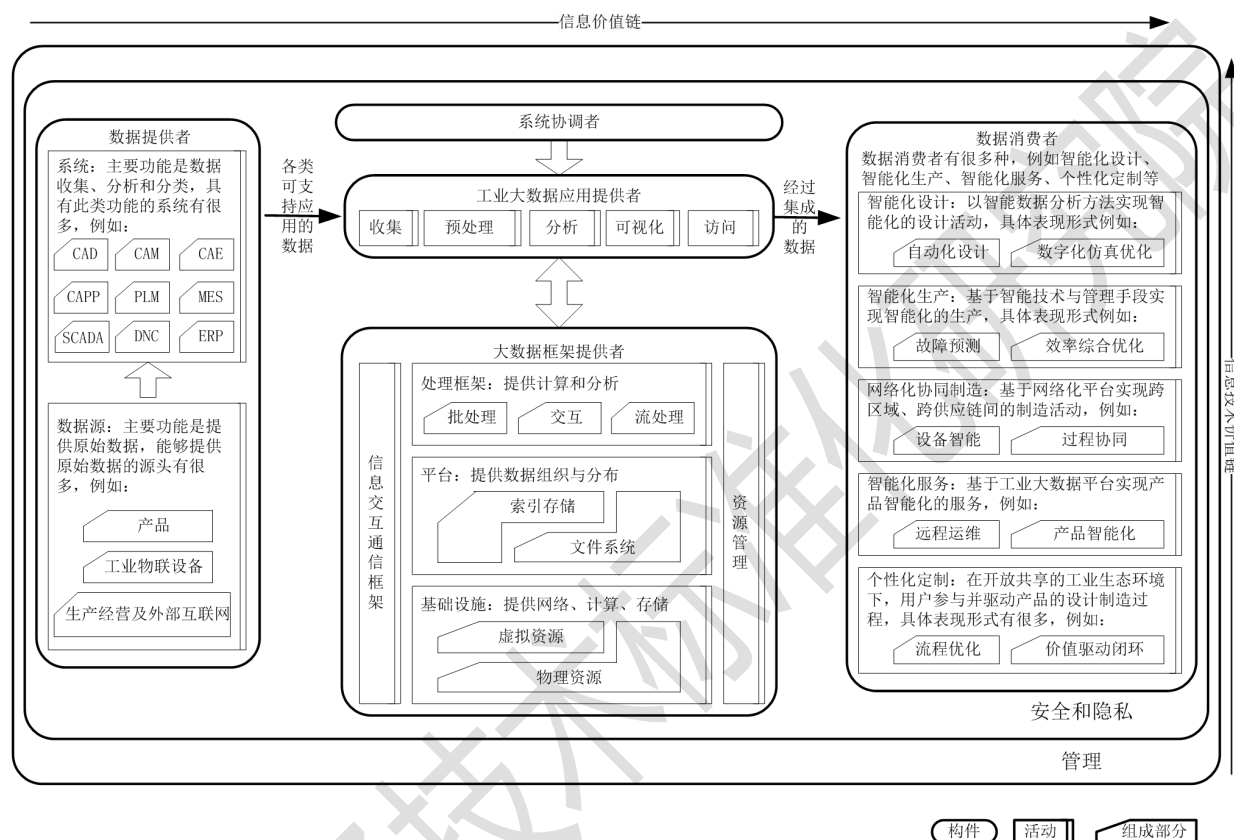


图 3-1 工业大数据应用参考架构

工业大数据应用参考架构将大数据参考架构的构件落实到了工业领域内的具体活动。工业大数据应用参考架构构件包括系统协调者、数据提供者、大数据应用提供者、大数据框架提供者、数据消费者、安全和隐私、管理。

### (1) 系统协调者

系统协调者的职责在于规范和集成各类所需的数据应用活动。系统协调者的职能包括配置和管理工业大数据应用参考架构中其它构件执行一个或多个工作负载，以确保各项工作能正常运行；为其它组件分配对应的物理或虚拟节点；对各组件的运行情况进行监控；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。

## (2) 数据提供者

数据提供者的基本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。数据提供者主要包括数据源和系统两部分，数据源是数据的产生处，它产生的数据通过信息系统收集、分析和分类后提供给工业大数据应用提供者；系统主要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。

## (3) 工业大数据应用提供者

工业大数据应用提供者的基本职能主要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。收集负责处理与数据提供者的接口和数据引入，根据工业大数据数据格式、类型的不同，通过引用对应的工业应用或构件，完成数据的识别和导入。预处理包括数据清洗、数据归约、标准化、格式化和存储。分析是指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。可视化是指对经处理、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。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。

## (4) 大数据框架提供者

大数据框架提供者的主要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。大数据框架提供者包括基础设施、平台、处理框架、信息交互/通信和资源管理 5 个活动。

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，包括网络、计算、存储、环境等。

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。处理框架通过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。

信息交互/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。

资源管理主要指计算、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。

#### (5) 数据消费者

数据消费者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等 5 种应用场景。本章第 3.3 节将对这 5 中典型应用场景进行介绍。

#### (6) 安全和隐私

在安全和隐私构件，是指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、软件和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。

#### (7) 管理

管理构件主要包括三方面功能，一是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、升级和补丁等功能。二是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，合理的分配和调度业务所需要的资源，做到自动化按需分配。三是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。

### 3.3 典型应用场景

随着信息化与工业化的深度融合，工业企业所拥有的数据也日益丰富，包括设计数据、传感数据、自动控制系统数据、生产数据、供应链数据等，数据驱动的价值体现及其带来的洞察力贯穿于智能制造生命周期的全过程。领先企业以平台为载体，不断形成针对制造业应用场景的大数据解决方案。制造和自动化领域的领军企业也依托长期积累的核心技术和行业知识，大力推广大数据在工业领域的应用，推动制造企业形成以数据驱动、快速迭代、持续优化的工业智能系统。面向制造业企业陆续形成的工业大数据平台正在为工业大数据在制造业的深入应用提供新技术、新业态和新模式。工业大数据已经成为工业企业生产力、竞争力、创新能力提升的关键，相关技术及产品已经逐步应用于工业企业和产业链的各环节，是驱动智能化产品、生产与服务，实现创新、优化的重要基础，体现在产品全生命周期中的各个阶段，正在加速工业企业的转型升级。近年来由智能制造、工业互联网发展催生的新模式、新应用，再次丰富了工业大数据的应用场景。

依据工业大数据支撑产品从订单到研发设计、采购、生产制造、交付、运维、报废、再制造的整个流程考虑，本白皮书将工业大数据典型的应用场景主要概括为智能化设计、智能化生产、网络化协同制造、智能化服务和个性化定制等五种模式，如图 3-2 所示。

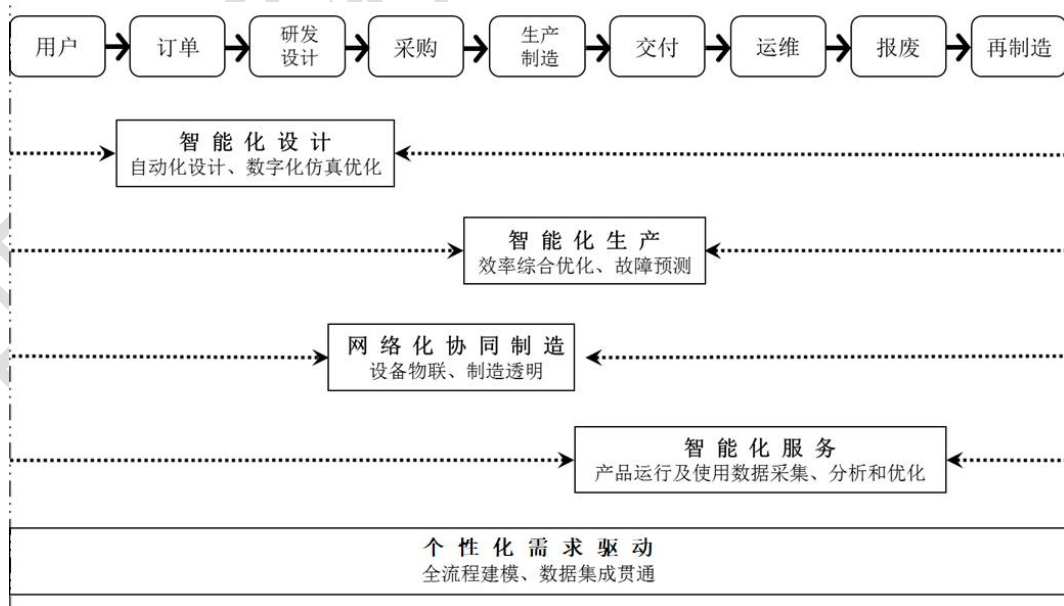


图 3-2 工业大数据典型应用场景



### 3.3.1 智能化设计

智能化设计是支撑工业企业实现全流程智能化生产的重要条件。设计数据包括企业设计人员或消费者借助各类辅助工具所设计的产品模型、个性化数据及相关资料，例如计算机辅助设计（CAD）、计算机辅助制造（CAM）、计算机辅助工程（CAE）、计算机辅助工艺设计（CAPP）、产品数据管理（PDM）等。工业大数据在设计环节的应用可以有效提高研发人员创新能力、研发效率和质量，推动协同设计。客户与工业企业之间的交互和交易行为将产生大量数据，挖掘和分析这些客户动态数据，能够帮助客户参与到产品的需求分析和产品设计等创新活动中，实现新型产品创新和协作的新模式。西门子在数字环境下构建基于模型和仿真的研发设计，有效提升了设计质量、节约研发成本；玛莎拉蒂通过数字化工具加速产品设计，实现开发效率提高 30%。

另一方面，传统的产品设计模式是基于设计师的灵感和经验，揣摩消费者的需求喜好，设计产品，针对性不强，不精确，工业大数据可以拉近消费者与设计师的距离，精准量化客户需求，指导设计过程，改变产品设计模式。通过将产品生命周期设计中各个环节所需要的知识资源有机的集成在一起，运用大数据相关技术，可以将产品生命周期设计所需大数据与各种设计过程集成，以高度有序化的方式展示产品生命周期大数据与设计的关系。GE 公司使用 Predix 平台助力自身发动机的设计优化，平台首先对产品交付后的使用数据进行采集分析，依托大量历史积累数据的分析和航线运营信息的反馈，对设计端模型、参数和制造端工艺、流程进行不断迭代优化，实现了发动机的设计改进和性能提升。金风科技通过工业大数据平台的大数据优化处理和针对载荷的定制化应用开发，将 5000 个工况的单轮仿真数据的后处理速度从半个月跑 1 轮，提速为一周跑 3 轮，单轮 30G 仿真数据的后处理速度从几个小时降低到了几分钟，极大地提升了风场定制化设计的迭代速度和开发效率。

工业大数据使产品生命周期大数据在设计过程中得到有效的应用、评价和推荐。设计知识能够快速地推送给所需要的人，并方便地融合员工在设计中产生的新知识，进一步丰富产品设计大数据。

### 3.3.2 智能化生产

智能化生产是新一代智能制造的主线，通过智能系统及设备升级改造及融合，促进制造过程自动化，流程智能化。从数据采集开始，生产阶段工业大数据的驱动力体现在数据关联分析和数据反馈指导生产。在生产阶段，对所采集的数据进行清洗、筛选、关联、融合、索引、挖掘，构建应用分析模式，实现数据到信息知识的有效转化。在制造阶段，通过对制造执行系统中所采集的生产单元分配、资源状态管理、产品跟踪管理等信息进行关联分析，为合理的库存管理、计划排程制定提供数据支撑；并且结合实时数据，对产品生产流程进行评估及预测，对生产过程进行实时监控、调整，并为发现的问题提供解决方案，实现全产业链的协同优化，完成数据由信息到价值的转变。工业大数据通过采集和汇聚设备运行数据、工艺参数、质量检测数据、物料配送数据和进度管理数据等生产现场数据，利用大数据技术分析和反馈并在制造工艺、生产流程、质量管理、设备维护、能耗管理等具体场景应用，实现生产过程的优化。

在制造工艺场景中，围绕生产工艺过程参数，设备运行状态参数与产品质量性能、生产线排产负荷、耗能等数据进行关联性深度挖掘，形成数据闭环，可得出工艺参数的最优区间、生产质量控制的最优调控手段等，提升产品制造品质。美林数据设计的图像实时采集与智能分析系统，能快速识别冲压件表面缺陷，同时关联了质检数据、生产过程工艺参数、产品设计参数，形成冲压产品质量问题分析管理的闭环连接，实现了冲压产品质量的精确控制和优化提升。阿里巴巴基于燃烧机理与锅炉运行数据，深度分析识别锅炉燃烧过程中的关键因子，找到了最优的锅炉操作参数，使燃煤效率提升了 4.1%，大大降低了能源消耗。

生产流程管理优化场景中，基于数据标准化思路的企业全流程的数据集成贯通及对生产进度、物料管理、企业管理等数据的分析，提升排产、进度、物料、人员等方面管理的准确性，企业的生产效率与管理水平显著提高。通过生产制造各个环节的数据整合集聚，对工业产品的生产过程数据进行关联，建立虚拟模型，仿真并优化生产流程。当所有流程和绩效数据都能在系统中重建时，对各环节制造数据的集成分析可以助力制造企业改进其生产流程。东方国信基于 BIOP 平台，通过工业大数据分析技术，建立虚实映射、实时监控、智能诊断、协同优化的数

字孪生，实现对工业实体生产制造过程和工艺流程的仿真及优化，在炼铁、工业锅炉、水电、空压机、能源等多个行业或领域落地应用。

质量管理场景中，基于产品质检数据和生产过程数据进行关联性分析，实现在线质量监测和质量异常与追溯分析，提升产品良品率。例如美林数据通过分析高压开关关键质检数据及过程加工数据，开展质量问题分类、质量问题原因分析、质量问题追溯等分析管控，促使产品良品率提升 1.3%，提升了企业的质量管理水平。联想通过工业大数据平台，提供面向制造、汽车、能源等行业的智能生产解决方案，为平台企业生产制造过程管理提供预见性的支撑与指导，提升精益管理水平。

能耗管理场景中，基于能耗数据的采集与分析，对设备、产线、场景能效使用进行合理规划，提高能源使用效率，实现节能减排。霍尼维尔通过能源管理系统提供能源设备管理、生产能耗分析、能源平衡管理等功能，帮助用户企业实现工艺流程的优化，设备改造升级，提升了能源利用率。

在复杂产线设备健康管理方面，随着科技的不断进步以及工业化水平的不断提高，工业系统的规模越来越大，集成的设备越来越多。系统内部通常存在复杂的耦合关系，其可靠性难以得到保障，一旦发生故障时，设备的停机损失将会非常大。当前对于复杂系统健康度与可靠性评估主要通过利用传感器对设备进行感知和从实时数据库系统获取设备振动、温度、压力、流量等数据，基于大数据平台对数据进行存储管理，借助人工智能算法对设备健康进行评估，实现设备故障预测和健康度监控。如美国电力公司（AEP）通过对变压器、断路器和蓄电池分别加装了 8600 个、11500 个和 400 多个传感器，基于其工业大数据平台，开展设备数据采集、诊断与分析，AEP 可以实时监控其设备参数、进行故障诊断预警，将设备寿命延长了 3 年，维护成本降低了 2.7%，设备维护效率提高了 4%，实现了设备预测性维护。

工业大数据助力解决生产过程复杂系统的精确建模、实时优化决策等关键问题，涌现出一批自学习、自感知、自适应、自控制的智能产线、智能车间和智能工厂，正在推动产品制造的高质、柔性、高效、安全与绿色，驱动生产过程的智能化升级。

### 3.3.3 网络化协同制造

在制造业向着大型、精密、数控、全自动趋势不断靠拢的时代下，基于工业大数据技术，将制造环节与设计、经销、运行、维护直至回收处理联系起来，由传统的数据孤岛转为信息化协同管理，推动产业链各环节的并行组织和协同优化。另一方面，借助大数据平台，将产业链各个环节的数据进行采集并输入到全生命周期数据库形成总知识库，通过信息技术、自动化技术、现代管理技术与制造技术相结合，构建面向企业的网络化协同制造系统，推动制造全产业链智能协同，优化生产要素配置和资源利用，消除低效中间环节，整体提升制造业发展水平和世界竞争力。工业大数据在网络化协同制造的应用主要体现在协同研发与制造、供应链管理体系优化、制造能力资源优化等方面。

协同研发与制造。基于设计资源的社会化共享和参与，企业能够立足自身研发需求开展众创、众包等研发新模式，提升企业利用社会化创新和资金资源能力。基于统一的设计平台与制造资源信息平台，集成设计工具库、模型库、知识库及制造企业闲置生产能力信息，产业链上下游企业可以实现多站点协同、多任务并行，加速新产品协同研发过程。河南航天基于航天云网 INDICS 平台，开展云端设计、建立涵盖复杂产品多学科专业的虚拟样机系统，实现了复杂产品的多学科设计优化、总体设计部与总装厂所的协同研发设计与工艺设计，研发周期缩短了 35%、资源利用率提升了 30%，生产效率提高了 40%。

供应链管理体系优化。追踪所有供应链中的在途部件是许多企业面临的最大挑战，通过空间地理数据的实时采集、实时数据的监控，提高供应链的可视化，实现供应链的透明化。通过空间数据的时空模式挖掘，为供应链物流的智能化运作提供实时决策依据，优化物流路径、优化排程调度，最后实现供应链物流的高效、智能化运行。通过全产业链大数据的整合，使整个生产系统达到协同优化，让生产系统更加动态灵活，进一步提高生产效率和降低生产成本，实现供应链配送体系优化和用户需求快速响应。通过对产品供应链的大数据进行分析，将带来仓储、配送、销售效率的大幅提升和成本的大幅下降。

制造能力资源优化。通过工业大数据技术的支撑，实现制造资源、制造能力、制造过程的信息透明，连通不同物理区域的多样化生产资源，根据订单需求完成

最优化的资源配置，实现高效高质的零部件协同制造，统一组装、交付，同时工业企业通过工业大数据平台可对外开放空闲制造能力，实现制造能力的在线租用和利益分配。例如沈阳机床基于工业大数据平台向奥邦锻造公司提供了沈机 i5 机床租赁服务，按照制造能力付费，有效降低了用户资金门槛，释放了产能。

以网络化协同制造为核心理念，大数据技术作为支撑，制造业企业内部及企业间在众多可靠的网络资源支持下实现了对不同产品各个阶段的增值，促进了创新资源、生产能力、市场需求的集聚与对接，提高了产业链上下游的资源整合能力，促进了全社会多元化制造资源的高度有效协同。

### 3.3.4 智能化服务

现代制造企业不再仅仅是产品提供商，而是提供产品、服务、支持、自我服务和知识的“集合体”。工业大数据与新一代技术的融合应用，赋予市场、销售、运营维护等产品全生命周期服务全新的内容，不断催生出制造业新模式、新业态，从大规模流水线生产转向规模化定制生产和从生产型制造向服务型制造转变，推动服务型制造业与生产性服务业大发展。

在市场营销环节，利用大数据挖掘用户需求和市场趋势，建立用户对商品需求的分析体系，挖掘用户深层次的需求，寻找机会产品，进行生产指导和后期市场营销分析；通过建立科学的商品生产方案分析系统，结合用户需求与产品生产，形成满足消费者预期的各品类生产方案等。航天云网基于 INDICS 平台的工业数据、供需服务信息，利用工业大数据关联和分析为企业提供产品推荐、销量预测、企业驾驶舱等数据服务，大大提升了平台上的企业营销水平及运营能力。

在售后服务环节，数据驱动企业创新服务模式，从被动服务、定期服务发展成为主动服务、实时服务。企业通过整合产品运行数据、销售、客户数据，将传统的诊断方法与基于知识的智能机械故障诊断方法相结合，结合设备状态监测技术、故障诊断技术和计算机网络技术，开展故障预警、远程监控、远程运维、质量诊断等大数据分析和预测，提供个性化、在线化、便捷化的智能化增值服务，形成“制造+服务”的新模式；建立产品电子履历，优化产品售后保障措施，实现产品全生命周期的质量信息追溯，提升用户满意度。三一重工打造的根云平台，

为行业企业提供基于物联网、大数据的云服务，当前平台已接入各类高价值设备 40 万台以上，采集近万个参数，连接数万亿资产，为客户开拓超百亿元的新业务，实现了制造企业的服务化转型。

### 3.3.5 个性化定制

个性化定制也是工业大数据应用的热点模式之一。通过工业大数据技术及解决方案，实现制造全流程数据集成贯通，构建千人千面的用户画像，并基于用户的动态需求，指导需求准确地转化为订单，满足用户的动态需求变化，最终形成基于数据驱动的工业大规模个性化定制新模式。

大规模个性化定制模式下，企业会提供一个互联网平台，作为与用户沟通交流的门户，在该平台上，消费者可以描述其个性化需求，进行个性化设计并下单，在收到产品后可提出意见与反馈，企业据此完善该用户的个性化数据，并进一步优化针对该用户的个性化设计。在大规模个性化定制生产中，数据起到了关键作用。需要采集客户个性化需求数据、工业企业生产数据、外部环境数据等信息，从而建立个性化产品模型，将产品方案、物料清单、工艺方案通过制造执行系统快速传递给生产现场，进行生产线调整和物料准备，快速生产出符合个性化需求的定制化产品。当用户的个性化需求订单产生的时刻，依据基于工业大数据构建的需求转化机制，可对制造过程中的变动做出快速整合和调整，柔性、动态地满足用户千人千面的个性化需求。

通过将工业大数据与大规模个性化定制模式的结合，形成支持工业产品开发个性化、设备管理个性化、企业管理个性化、人员管理个性化、垂直行业个性化等一系列满足用户个性化需求的工业价值创造新模式，为工业企业显著降低成本，形成价值创造的新动能。例如联想自 2011 年起实施了全集团的数字化转型和全球化的大规模个性化定制实践，构建了大规模个性化定制的先进计算与存储设备制造支撑平台，形成了覆盖从用户需求、产品设计、研发、柔性生产制造、供应链和金融等关键环节的全价值链精准赋能，实现了企业内部不同产品线和外部关联企业间的工业领域知识和数据互通，个性化定制生产和生产能力分享，显著降低了企业的生产经营成本、缩短产品升级周期、提高产品生产效率。

### 3.4 产业发展分析

工业是我国大数据与实体经济深度融合的主战场。近年来，随着工业互联网的快速发展及两化融合深入推进，我国工业大数据发展呈现爆发式增长态势，应用广度和应用深度不断加强，对工业转型升级具有重要支撑引领作用。

从供给侧看，工业大数据供给侧能力持续提升，涌现出一批专精特新企业，成为推动我国工业大数据发展的中坚力量。一是由传统工业制造企业数字化、软件化、平台化发展，出现了一批具有较强数据汇聚能力的衍生型企业，如航天云网、树根互联、石化盈科等；二是软件企业向工业领域渗透，出现的技术型企业，如昆仑智汇、东方国信等企业在工业数据建模、分析处理等领域不断突破核心技术；三是互联网企业积极进入工业领域，如阿里推出“ET 工业大脑”等产品和服务，腾讯推出工业互联网“木星云”平台。从需求侧看，随着智能制造、工业互联网等国家战略的逐一推进，个性化定制、网络化延伸以及智能化设计、生产、服务等新模式不断出现，对于工业大数据技术、产品、平台的需求量不断增大，为工业大数据提供了充足的应用场景。

然而，在我国工业大数据产业发展不断优化提升的同时，仍需要清楚认识到我国工业大数据仍存在物联数据无法获取、格式不统一，数据产权不清晰、数据壁垒难以打破，全产业链数据应用不足等问题。主要原因在于，第一，在我国国产工业软件、高端物联设备核心技术供给不足，而国外设备读写不开放，数据无法读取或者格式多样，无法直接利用；第二，面对体量大、分布广、结构复杂、类型多样化的工业数据，目前工业行业整体数据资源管理水平不足，难以管理企业内部和外部各类数据，更无法充分分析和利用；第三，缺乏可用、好用、可信的工业大数据平台，难以充分利用工业全产业链上下游的数据，以实现人、机、物等各类工业要素、工业业务流程以及产业链上下游企业间更大范围的实施链接与智能交互，推动工业生产的资源优化、协同制造和服务延伸。

## 4 工业大数据关键技术

### 4.1 工业大数据技术架构

围绕工业大数据的全生命周期，本白皮书提出了工业大数据技术参考架构如图 4-1 所示。

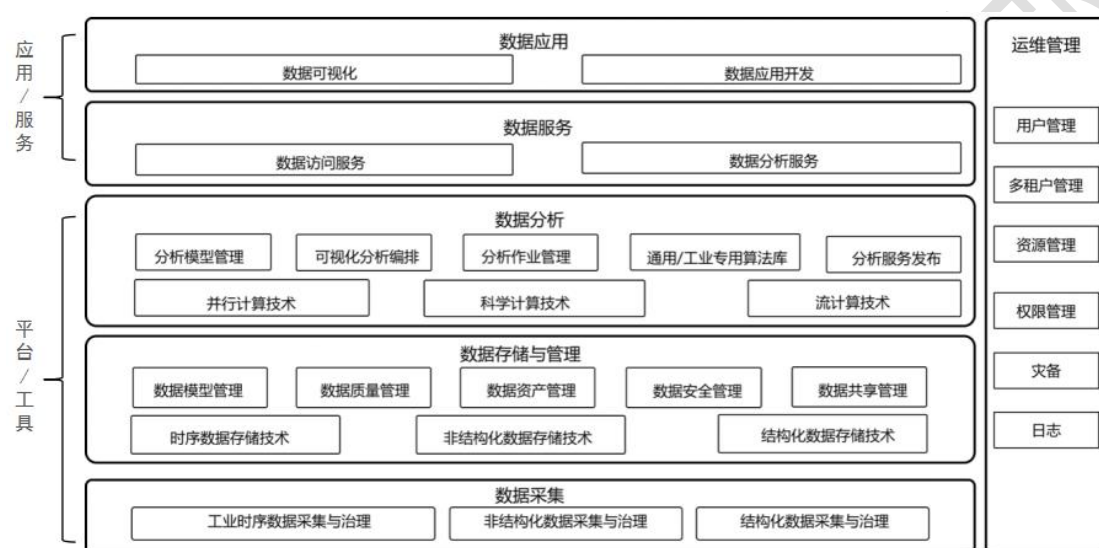


图 4-1 工业大数据技术参考架构

工业大数据技术参考架构以工业大数据的全生命周期为主线，从纵向维度分为平台/工具域和应用/服务域。平台/工具域主要面向工业大数据采集、存储管理、分析等关键技术，提供多源、异构、高通量、强机理的工业大数据核心技术支撑；应用/服务域则基于平台域提供的技术支撑，面向智能化设计、网络化协同、智能化生产、智能化服务、个性化定制等多场景，通过可视化、应用开发等方式，满足用户应用和服务需求，形成价值变现。

工业大数据技术参考架构从技术层级上具体划分如下：

数据采集层，包括时序数据采集与治理、结构化数据采集与治理和非结构化数据采集与实时处理。海量工业时序数据具有 7\*24 小时持续发送，存在峰值和滞后等波动，质量问题突出等特点。需要构建前置性数据治理组件与高性能时序数据采集系统。针对结构化与非结构化数据，需要构建同时兼顾可扩展性和处理性能的数据采集系统。数据采集层的数据源主要包括通过 ETL 方式同步的企业生



产经营相关的业务数据、实时或批量采集的设备物联数据和从外部获取的第三方数据。

数据存储与管理层，包括大数据存储技术和管理功能。利用大数据分布式存储的技术，构建在性能和容量都能线性扩展的时序数据存储、结构化数据存储和非结构化数据存储等。基于以上存储技术并结合工业大数据在数据建模、资产沉淀、开放共享等方面的特殊需求，构建数据模型管理、数据质量管理、数据资产管理、数据安全管理和数据共享管理技术体系。

数据分析层，包括基础大数据计算技术和大数据分析服务功能，其中基础大数据计算技术包括并行计算技术、流计算技术和数据科学计算技术。在此之上构建完善的大数据分析服务功能来管理和调度工业大数据分析，通过数据建模、数据计算、数据分析形成知识积累，以实现工业大数据面向生产过程智能化、产品智能化、新业态新模式智能化、管理智能化以及服务智能化等领域的数据分析。大数据分析服务功能包括分析模型管理、可视化编排、分析作业管理、工业专用/通用算法库和分析服务发布。

数据服务层是利用工业大数据技术对外提供服务的功能层。包括数据访问服务和数据分析服务。其中数据访问服务对外提供大数据平台内所有原始数据、加工数据和分析结果数据的服务化访问接口和功能；数据分析服务对外提供大数据平台上积累的实时流处理模型、机理模型、统计模型和机器学习模型的服务化接口。数据服务层提供平台各类数据源与外界系统和应用程序的访问共享接口，其目标是实现工业大数据平台的各类原始、加工和分析结果数据与数据应用和外部系统的对接集成。

数据应用层，主要面向工业大数据的应用技术，包括数据可视化技术和数据应用开发技术。综合原始数据、加工数据和分析结果数据，通过可视化技术，将多来源、多层次、多维度数据以更为直观简洁的方式展示出来，易于用户理解分析，提高决策效率。综合利用微服务开发框架和移动应用开发工具等，基于工业大数据管理、分析技术快速实现工业大数据应用的开发与迭代，构建面向实际业务需求的，数据驱动的工业大数据应用，实现提质降本与增效。数据应用层通过生成可视化、告警、预测决策、控制等不同的应用，从而实现智能化设计、智能

化生产、网络化协同制造、智能化服务和个性化定制等典型的智能制造模式，并将结果以规范化数据形式存储下来，最终构成从生产物联设备层级到控制系统层级、车间生产管理层级、企业经营层级、产业链上企业协同运营管理的持续优化闭环。

此外运维管理层也是工业大数据技术参考架构的重要组成，贯穿从数据采集到最终服务应用的全环节，为整个体系提供管理支撑和安全保障。

围绕工业大数据技术架构，本章将针对工业大数据平台及其关键支撑技术展开具体介绍。

## 4.2 工业大数据平台

工业大数据平台是工业大数据技术具体应用的载体，是推进工业大数据技术深度应用、提升工业大数据在产业中整体发展水平的重要基石。从企业个体角度来看，工业大数据平台是整个企业工业大数据应用的核心。一方面平台通过提供数据采集接口，对企业经营管理的业务数据、机器设备互联数据以及销售运维等外部数据进行采集、清洗，并基于工业大数据处理、分析、建模等关键技术，根据具体应用场景及需求，结合领域知识和算法，实现顶层应用支撑，产生应用价值。从全产业链条角度看，工业大数据平台是产业链实现数据互联的重要枢纽。基于构建安全、可信的数据交换机制，打通不同企业的工业大数据平台，才能在商业生态系统中进行安全的数据交换和方便的数据连接，对接工业制造和智能服务，实现全产业链条的协同制造。

从平台建设角度看，工业大数据平台首先需要大数据基础设施的支撑，比如基础的计算、存储、网络设备，云数据中心，云计算平台等；其次需要工业数据采集、存储、分析、可视化等关键技术支撑，以满足在具有多模态、高通量、强关联特征的工业大数据环境中支撑工业应用；此外还需具备运营管理保障，通过监控、告警、备份、恢复和优化等方面，保障整个数据处理架构的稳定高效运营。

从平台功能角度看，工业大数据平台实质上是大数据平台在工业领域的应用，因此工业大数据平台应当具备通用大数据平台针对数据全生命周期处理的所有功能。但由于工业数据的复杂来源，工业大数据平台既要处理来自业务系统的

工程数据和管理数据，又包括来自物联网的高频时序数据和来自互联网的资源数据，工业大数据平台需要加强对不同来源数据的采集、集成、分析能力以及针对工业环境下的安全管控功能。

目前我国部分企业已经具备自主研发的工业大数据平台，在工业大数据平台的工业大数据采集、工业大数据存存储管理、工业大数据分析关键支撑技术上也已经有所突破，然而针对工业大数据平台的边界还未在业内达成统一，需要标准化的手段进行规范，支撑我国工业大数据平台的发展。本章后续将聚焦工业大数据平台的关键支撑技术展开介绍。

### 4.3 工业大数据采集技术

数据采集方面，以传感器为主要采集工具，结合 RFID、条码扫描器、生产和监测设备、PDA、人机交互、智能终端等手段采集制造领域多源、异构数据信息，并通过互联网或现场总线等技术实现原始数据的实时准确传输。工业大数据分析往往需要更精细化的数据，对于数据采集能力有着较高的要求。例如高速旋转设备的故障诊断需要分析高达每秒千次采样的数据，要求无损全时采集数据。通过故障容错和高可用架构，即使在部分网络、机器故障的情况下，仍保证数据的完整性，杜绝数据丢失。同时还需要在数据采集过程中自动进行数据实时处理，例如：校验数据类型和格式，异常数据分类隔离、提取和告警等。

数据采集是获得有效数据的重要途径，是工业大数据分析和应用的基础。数据采集与治理的目标是从企业内部和外部等数据源获取各种类型的数据，并围绕数据的使用，建立数据标准规范和管理机制流程，保证数据质量，提高数据管控水平。

工业大数据的采集主要是通过 PLC、SCADA、DCS 等系统从机器设备实时采集数据，也可以通过数据交换接口从实时数据库等系统以透传或批量同步的方式获取物联网数据。同时还需要从业务系统的关系型数据库、文件系统中采集所需的结构化与非结构化业务数据。针对海量工业设备产生的时序数据，如设备传感器指标数据、自动化控制数据，需要面向高吞吐、7\*24 小时持续发送，且可容忍峰值和滞后等波动的高性能时序数据采集系统。针对结构化与非结构化数据，需

要同时兼顾可扩展性和处理性能的实时数据同步接口与传输引擎。针对仿真过程数据等非结构化数据具有文件结构不固定、文件数量巨大的特点，需要元数据自动提取与局部性优化存储策略，面向读、写性能优化的非结构化数据采集系统。

在工业大数据中，数据质量问题一直是许多企业所面临的挑战，这主要受制于工业环境中数据获取手段的限制，包括传感器、数采硬件模块、通信协议和组态软件等多个技术限制，因此需要在实时数据过程中对数据质量进行监测、分析和处理，在源头尽可能地消除问题。针对工业时序数据质量问题，如数据格式不规范、错漏字段、命名版本管理缺失等问题，需要前置性数据治理模块对数据进行实时处理，通过实时规则与模式匹配逐条核查时序数据的质量，建立后效性多变量关联的机理约束模型来检测深层次数据质量问题。

## 4.4 工业大数据存储与管理技术

工业大数据存储与管理技术是针对工业大数据具有多样性、多模态、高通量和强关联等特性，研发的面向高吞吐量存储、数据压缩、数据索引、查询优化和数据缓存等能力的关键技术，更多的体系化数据管理方法将在第5章展开介绍，本节主要围绕工业大数据存储与管理技术中多源异构数据高效管理技术和多模态数据集成技术两类关键技术进行介绍。

### 4.4.1 多源异构数据管理技术

多源异构数据是指数据源不同、数据结构或类型不同的数据集合。各种工业场景中存在大量多源异构数据例如，在诊断设备故障时，通过时间序列数据可以观测设备的实时运行情况；通过BOM图数据可以追溯出设备的制造情况，从而发现是哪些零部件问题导致异常运行情况；通过非结构化数据可以有效管理设备故障时的现场照片、维修工单等数据；键值对数据作为灵活补充，能方便地记录一些需要快速检索的信息。

数据源不同、数据类型不同，使得这类数据集的使用变得非常复杂，因此大规模多源异构数据管理技术变得十分重要。为使这些多源异构数据各自发挥其价值，不仅需要高效的存储管理优化与异构的存储引擎，在此基础上还需要能够通

过数据融合对数据的元数据定义和高效查询与读取进行优化,实现多源异构数据的一体化管理,从而最大程度上榨取数据价值。

多源异构数据管理需要突破的是针对不同类型数据的存储与查询技术,并在充分考虑多源异构数据的来源和结构随着时间推移不断增加与变化的特定的情况下,研究如何形成可扩展的一体化管理系统。

多源异构数据管理需要从系统角度,针对工业领域涉及的数据在不同阶段、不同流程呈现多种模态(关系、图、键值、时序、非结构化)的特点,研制不同的数据管理引擎致力于对多源异构数据进行高效地采集、存储和管理。当前,国产数据库及数据管理引擎仍处于新兴发展阶段,在传统的结构化数据之外,针对多源异构数据(包括时序数据、过程与BOM图数据,以及工程非结构化数据等),开发稳定而高效的数据管理引擎、并真正落地到工业领域变得愈发重要。

针对海量的工业时序数据在查询高效性和接入吞吐量方面的需求,需要构建能够满足数据边缘接入与缓存、高性能读写、高效率存储、查询与分布式分析一体化的时序数据管理系统,配合缓存、分布式计算与存储框架等组件,以满足功能和易用性需求。同时需要提供基于SQL标准的数据查询接口给工业用户以降低使用门槛。

工业领域的非结构化数据,面向仿真、试验等场景的海量小文件的挑战,要求按产品生命周期、BOM结构等多种维度进行灵活组织和高效查询,同时对数据能够进行批量读取分析,因此需要构建面向工业场景的支持海量非结构化文件建模、存储、查询和读取的技术系统。

多源异构数据管理技术可有效解决大数据管理系统中由模块耦合紧密、开放性差而导致的系统对数据多样性和应用多样性的适应能力差的问题,使大数据管理系统能够更好地适应数据和应用的多样性并能够充分利用开源软件领域强大的技术开发和创新能力。针对企业自身数据类型和特点,通过量体裁衣式的构件组合,能够帮助工业企业快速开发和定制适合自身需求的制造业大数据管理系统。

#### 4.4.2 多模态数据集成技术

工业大数据来源十分广泛，包括但不限于研发环节的非结构化工程数据、传统的企业信息管理系统、服务维修数据和产品服役过程中产生的机器数据等。这些数据格式异构、语义复杂且版本多变。在工业大数据应用中，希望能够将多模态数据有机地结合在一起，发挥出单一模态数据无法挖掘出的价值。

数据集成是将存储在不同物理存储引擎上的数据连接在一次，并为用户提供统一的数据视图。传统的数据集成领域中认为，由于信息系统的建设是阶段性和分布性的，会导致“信息孤岛”现象的存在。“信息孤岛”造成系统中存在大量冗余数据，无法保证数据的一致性，从而降低信息的利用效率和利用率，因此需要数据集成。在工业大数据中，重点不是解决冗余数据问题，而更关心数据之间是否存在某些内在联系，从而使得这些数据能够被协同地用于描述或者解释某些工业制造或者设备使用的现象。

数据集成的核心任务是要将互相关联的多模态数据集成到一起，使用户能够以透明的方式访问这些数据源。集成是指维护数据源整体上的数据一致性、提高信息共享利用的效率；透明的方式是指用户无需关心如何实现对异构数据源数据的访问，只关心以何种方式访问何种数据。更进一步地，数据融合是在数据集成的基础上，刻画出不同数据之间的内在联系，并允许用户根据这些内在联系进行数据查询。

在数据生命周期管理中，多模态数据存储分散、关系复杂，在研发、制造周期以 BOM 为主线，在制造、服务周期以设备实例为中心，BOM 和设备的语义贯穿了工业大数据的整个生命周期。因此，以 BOM 和设备为核心建立数据关联，可以使得产品生命周期的数据既能正向传递又能反向传递，形成信息闭环，而对这些多模态数据的集成是形成数据生命周期信息闭环的基础。

针对工业领域在研发、制造和服务各个周期产生的多模态数据，如核心工艺参数、检测数据、设备监测数据等，及其存储分散、关系复杂的现状，需要实现统一数据建模，定义数字与物理对象模型，完成底层数据模型到对象模型映射。在多模态数据集成模型的基础上，根据物料、设备及其关联关系，按照分析、管

理的业务语义，实现多模态数据的一体化查询、多维分析，构建虚实映射的全生命周期数据融合模型。在多模态数据集成模型基础上，针对多模态数据在语义与数据类型上的复杂性，实现语义模糊匹配技术的异构数据一体化查询。

多模态数据集成技术目前仍面临如下两个难点：

(1) 语义融合：基于多层面特征的多模态数据融合方法以及支持制造业多模态数据的统一数字物理对象模型，实现多模态数据与统一数字物理对象之间的映射与数据的统一建模和管理。基于制造业的全生命周期，如何建立基于制造语义、业务机理、制造流程以及 BOM 结构关系的人、机、料、法、环的数据语义融合方法，挖掘数据之间的潜在关联关系，完成在数据的语义关联与融合。

(2) 一体化查询优化：在统一查询执行和优化引擎中设计协同算法，由统一查询执行和优化引擎调用各模态数据查询引擎协同完成全局查询优化，从而在混合计算与网络环境下保证查询执行正确性、数据一致性以及性能最优。同时考虑在各种实时、准实时、非实时的多模态数据查询混合并发时负载不均的问题，建立一套多模态数据管理引擎的负载均衡理论，最终实现不同查询引擎之间的负载均衡。

多模态数据集成技术将引发出一系列与产品创新与智能服务相关的数据管理问题及其可能的解决方案，为进一步研究新一代信息物理系统的数据管理与服务技术开辟空间。

## 4.5 工业大数据分析技术

工业大数据具有实时性高、数据量大、密度低、数据源异构性强等特点，这导致工业大数据的分析不同于其他领域的大数据分析，通用的数据分析技术往往不能解决特定工业场景的业务问题。工业过程要求工业分析模型的精度高、可靠性高、因果关系强，这样才能满足日常工业生产需要，而纯数据驱动的数据分析手段往往不能达到工业场景的要求。工业数据的分析需要融合工业机理模型，以“数据驱动+机理驱动”的双驱动模式来进行工业大数据的分析，从而建立高精度、高可靠性的模型来真正解决实际的工业问题。因此，工业大数据分析的特征是强调专业领域知识和数据挖掘的深度融合。本节主要对时序模式分析技术、工

业知识图谱技术、多源数据融合分析技术等三种典型的工业大数据分析技术进行介绍。

#### 4.5.1 时序模式分析技术

伴随着工业技术的发展，工业企业的生产加工设备、动力能源设备、运输交通设备、信息保障设备、运维管控设备上都加装了大量的传感器，如温度传感器、振动传感器、压力传感器、位移传感器、重量传感器等，这些传感器在不断产生海量的时序数据，提供了设备的温度、压力、位移、速度、湿度、光线、气体等信息。对这些设备传感器时序数据分析，可实现设备故障预警和诊断、利用率分析、能耗优化、生产监控等。但传感器数据的很多重要信息是隐藏在时序模式结构中，只有挖掘出背后的结构模式，才能构建一个效果稳定的数据模型。

工时序数据的时间序列类算法主要分六个方面：时间序列的预测算法如 ARIMA, GARCH 等；时间序列的异常变动模式检测算法，包含基于统计的方法、基于滑动窗窗口的方法等；时间序列的分类算法，包括 SAX 算法、基于相似度的方法等；时间序列的分解算法，包括时间序列的趋势特征分解、季节特征分解、周期性分解等；时间序列的频繁模式挖掘，典型时序模式智能匹配算法（精准匹配、保形匹配、仿射匹配等），包括 MEON 算法、基于 motif 的挖掘方法等；时间序列的切片算法，包括 AutoPlait 算法、HOD-1D 算法等。

工业大数据分析的一个重要应用方向是对机器设备的故障预警和故障诊断，其中设备的振动分析是故障诊断的重要手段。设备的振动分析需要融合设备机理模型和数据挖掘技术，针对旋转设备的振动分析类算法主要分成三类：振动数据的时域分析算法，主要提取设备振动的时域特征，如峭度、斜度、峰度系数等；振动数据的频域分析算法，主要从频域的角度提取设备的振动特征，包括高阶谱算法、全息谱算法、倒谱算法、相干谱算法、特征模式分解等；振动数据的时频分析算法，综合时域信息和频域信息一种分析手段，对设备的故障模型有较好的提取效果，主要有短时傅里叶变换、小波分析等。



## 4.5.2 工业知识图谱技术

工业生产过程中会积累大量的日志文本，如维修工单、工艺流程文件、故障记录等，此类非结构化数据中蕴含着丰富的专家经验，利用文本分析的技术能够实现事件实体和类型提取（故障类型抽取）、事件线索抽取（故障现象、征兆、排查路线、结果分析），通过专家知识的沉淀实现专家知识库（故障排查知识库、运维检修知识库、设备操作知识库）。

针对文本这类的非结构化数据，数据分析领域已经形成了成熟的通用文本挖掘类算法，包括分词算法（POS tagging、实体识别）、关键词提取算法（TD-IDF）、词向量转换算法、词性标注算法（CLAWS、VOLSUNGA）、主题模型算法（如 LDA）等。但在工业场景中，这些通用的文本分析算法，由于缺乏行业专有名词（专业术语、厂商、产品型号、量纲等）、语境上下文（包括典型工况描述、故障现象等），分析效果欠佳。这就需要构建特定领域的行业知识图谱（即工业知识图谱），并将工业知识图谱与结构化数据图语义模型融合，实现更加灵活的查询和一定程度上的推理。

## 4.5.3 多源数据融合分析技术

在企业生产经营、营销推广、采购运输等环节中，会有大量的管理经营数据，其中包含着众多不同来源的结构化和非结构化数据，例如来源于企业内部信息系统（CRM、MES、ERP、SEM）的生产数据、管理数据、销售数据等，来源于企业外部的物流数据、行业数据、政府数据等。利用这些数据可实现市场洞察、价格预测、供应链协同、精准销售、市场调度、产品追溯、能力分析、质量管控等等。通过对这些数据的分析，能够极大的提高企业的生产加工能力、质量监控能力、企业运营能力、市场营销能力、风险感知能力等。但多源数据也带来一定的技术挑战，不同数据源的数据质量和可信度存在差异，并且在不同业务场景下的表征能力不同。这就需要一些技术手段去有效融合多源数据。

针对多源数据分析的技术主要包括：统计分析算法、深度学习算法、回归算法、分类算法、聚类算法、关联规则等。可以通过不同的算法对不同的数据源进

行独立的分析，并通过对多个分析结果的统计决策或人工辅助决策，实现多源融合分析。也可以从分析方法上实现融合，例如通过非结构化文本数据语义融合构建具有制造语义的知识图谱，完成其他类型数据的实体和语义标注，通过图模型从语义标注中找出跨领域本体相互间的关联性，可以用于识别和发现工业时序数据中时间序列片段对应的文本数据（维修报告）上的故障信息，实现对时间序列的分类决策。

## 4.6 工业大数据的前沿技术趋势

目前深度学习、知识图谱、虚拟现实等前沿技术已经在互联网环境得到了广泛的应用，但是这些前沿技术如何在工业场景中发挥其价值，还需要进一步的探索。

在深度学习应用方面，利用深度学习算法在工业场景中开展图像和视频处理将成为重要发展方向。如管道焊缝 X 光片的缺陷识别，基于深度学习的 X 射线焊缝图像识别可以对缺陷的特征进行提取和自动研判；打造智能安全工厂，现有的生产工厂的安全生产监控系统，人工监控容易疏忽，仅用作事后取证的记录，没有发挥安全预警的价值，基于视频跟踪的深度学习技术，实时监控工厂安全状态，对非安全行为进行实时预警。

在工业知识图谱方面，如何在工业生产过程中提取工业语义关键信息并关联形成具备专业特点的工业知识图谱是下一步探索重点。一个值得探索的重要方向是围绕复杂装备运维服务阶段大量自然语言工单数据，利用复杂装备设计研发阶段形成专业词库，提取面向复杂装备具有的工业语义关键信息，并自动将这些关键语义进行关联形成具有专业特点的工业知识图谱。具体包括事件实体和类型提取（比如故障类型抽取）、事件线索抽取（比如故障现象、征兆、排查路线、结果分析），并将知识图谱与设备资产档案（时序数据）进行关联，实现工况上下文中的知识推理。

虚拟现实作为智能制造的核心技术之一，目前主要探索方向是通过数字孪生体，实现物理世界到虚拟世界的映射。基于数字孪生体技术，可以实现工业生产制造过程中产品设计的协同化、远程运维的智能化、产品试验完全仿真化，大幅度提升工业生产效率和智能水平。

## 5 工业大数据管理

### 5.1 数据管理体系

2018年3月，在工信部、国标委指导下，全国信标委大数据标准工作组组织编写的（GB/T 36073-2018）《数据管理能力成熟度评估模型》国家标准（DCMM）正式发布。作为我国首个数据管理领域国家标准，《数据管理能力成熟度评估模型》国家标准在编制过程中借鉴了国内外数据管理相关理论思想，并充分结合我国大数据行业发展的趋势，提出了基于我国产业发展现状的、针对企业的数据管理能力成熟度评估模型，将组织数据管理能力划分为8个能力域：数据战略、数据治理、数据架构、数据标准、数据质量、数据安全、数据应用、数据生存周期，如图5-1所示。



图 5-1 DCMM 能力域

为了更好的推进国家标准的落地应用，规范和指导相关组织提升数据管理水平，全国信标委大数据标准工作组在全国范围内积极开展《数据管理能力成熟度评估模型》国家标准试点示范工作，涵盖金融、能源、互联网、工业等多个领域的30余家企事业单位，取得了良好的效果，其中包括工业企业7家。

随着新一代信息技术和实体经济的深度融合，工业企业在生产制造过程中不断积累生产、研发、经营管理、运维等数据，累计数据体量巨大且种类繁多。但由于工业领域信息化建设相对落后，数据管理机制缺失、以及现场工况恶劣、缺

乏过程控制机制、安全防护不到位等原因，导致工业领域数据管理水平普遍不足。结合《数据管理能力成熟度评估模型》国家标准试验验证工作在工业企业应用现状，当前工业企业着重在数据资源梳理（数据资源编目）、数据质量管理、主数据管理、数据安全等方面亟需提升。

## 5.2 数据资源编目

工业企业的决策者们，大都面临这样一个问题，不明确企业内部都有哪些数据，数据存储在哪里，如何才能更好的发现数据、管理数据，并充分挖掘其价值。这一问题可以通过数据资源编目解决。

数据资源编目是通过对数据资源依据规范的元数据描述，对企业数据资产进行逻辑集中的一种方式。通过编目形成的数据资源目录含有各数据资源的描述信息，便于用户对数据资源的检索、定位和获取，帮助企业决策者们明晰企业所拥有的数据资产以及其数据资产的管理状态，提供企业数据资产显性化的应用入口，真正实现数据的可见、可管、可用。本节将从数据资源目录应用架构和数据资源目录建设两方面展开，为工业企业数据资源编目提供依据。

### 5.2.1 数据资源目录应用架构

数据资源目录应用参考架构如图 5-2 所示。

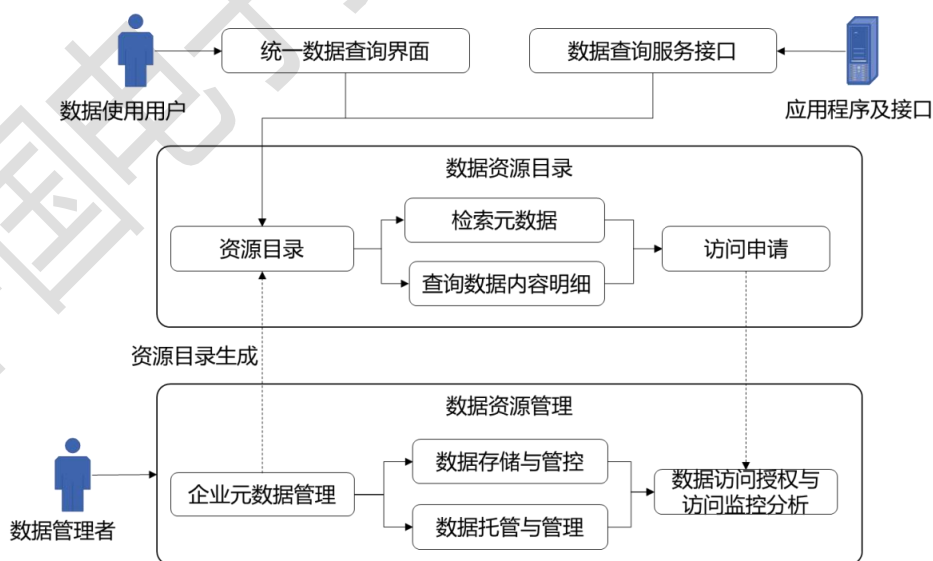


图 5-2 数据资源目录应用架构

数据资源目录应用包括对企业内部数据资源的统一管理和基于数据资源目录的对外服务两部分。

对企业内部的统一管理，主要是数据管理者对企业元数据进行统一的标准化的管理，并基于元数据实现企业数据的统一归集（主要是数据管控、数据托管两种方式），进而自动生成企业数据资源目录；并控制数据对服务的权限范围、数据质量等。

基于数据资源目录的对外服务，主要面向企业数据使用方进行企业数据的访问、获取等，包括用户对元数据的统一检索、以及对数据的查询服务等，对外的数据服务包括使用者直接登录平台进行数据访问、第三方系统通过接口等方式进行数据获取等。各种访问方式均受平台统一的权限控制，需要进行访问申请。

## 5.2.2 数据资源编目

企业数据资源编目主要包括以下工作：

(1) 企业元数据统一管理：根据不同的数据管理需求，进行不同方式的元数据注册及管理，对于存储及管控的数据，需开展标准设计及元数据创建工作；对于托管及存储的数据，需开展已有元数据的自动解析工作，获取元数据信息。

(2) 企业数据统一归集：基于统一管理的元数据进行企业数据的统一管理，实现企业数据的逻辑中心建设，不同的数据管理模式有不同的数据归集方式。管控型数据，拥有数据的增删权限，能够实现数据质量、维护规则等的强管控；托管型数据，通常由第三方业务系统管理，但是可以作为企业统一的数据资源对外进行共享的数据；接口型数据，不以数据实体存在，而且某些第三方系统发布的可访问其数据的接口，也需进行统一管理；非结构化数据，需纳入数据资产形成统一的资产目录，可对外范围并进行预览下载等。

(3) 自动生成企业数据资源目录：基于统一管理的元数据、统一归集的数据自动生成企业数据资产目录，展现数据有哪些、是什么、谁管理、在哪里等信息，并且提供基于数据模型、属性字段、类型、来源、管理部门等信息的查询与检索。当企业有新的数据资源注册到数据平台时，数据资源目录自动更新。

(4) 数据资源目录对外提供服务：用户可直接通过数据资产目录对企业元数据、企业数据明细进行访问，访问的方式同样提供多种，包括登录系统进行页面访问数据导出等、调用本地接口获取数据、通过 API 对外提供数据等。

(5) 数据服务权限控制：以数据资源目录为入口进行数据访问将控制各方式的数据访问权限，包括表范围、字段范围、数据量范围等；并需对访问情况进行监控，以实时掌握数据资产的应用情况及其产生的价值。

## 5.3 数据质量管理

工业大数据覆盖了产品研发、生产、市场、客户、物流供应链、售后服务、财务、人力、生产设备仪器仪表、传感器、产品、环境法规、社会经济等数据，涵盖流程长、种类多、范围广。数据质量的问题往往表现为一个或多个质量维度存在缺失从而无法满足企业和用户的实际需要。主要表现在准确性、完整性、一致性、及时性、唯一性等方面。数据质量管理作为智能制造顶层设计、过程控制和事后评价的重要依据，如何建立客观、全面、可行的数据质量检测、治理与持续改善机制，是大数据是否能够良好服务于智能制造的新的挑战。

### 5.3.1 工业大数据质量问题

工业大数据从应用场景上可划分为 IT 域及 OT 域数据。OT 域的数据主要来自于工业现场的物联网络、生产制造装备和基础自动化系统等采集的数据，主要由时序数据和非时序数据两大类组成，时序数据包括描述设备运行过程的温度、压力、振动、流量等运行监测数据以及描述产品所处阶段的声、光、电数据；非时序数据，主要来源于工业系统生成的日志数据以及工业生产过程积累的经验知识等。工业 IT 域数据主要是 MES、WMS、SCM 等产生的生产制造数据，ERP、MRP 等产生的企业运营数据以及通过企业外网获取的互联网数据在内的结构化与半结构化数据。

从来源层面，工业大数据向下涵盖了 OT 域的人、设备、物带来的物联数据；向上涵盖了 IT 域的制造执行、企业管理、客户关系管理等企业信息化数据，随着智能制造对数字化的依赖性持续增加，由于器件和系统故障、现场恶劣工况、

违反工艺规程、调度管理不当、数据格式不一致、数据字段重复或缺失等原因导致的数据质量问题，更加显著地影响着工业过程和制成品的品质。工业大数据的质量问题主要表现在以下几个方面：

数据失真和失准。受制于工业现场的一系列恶劣工况，在工业 OT 域，现场物联网络、生产制造装备、过程控制设备均不同程度地存在数据失真。由于工业现场的高温、高压、强振动、大噪声、大粉尘、磁场变化大等环境因素的影响，可能造成现场物联网络采集到的声、光、电、RFID 等各类工业运行数据出现误采、漏采、丢包等数据失真问题。由于工业现场管理规程及管理行为的不健全和不规范，工业生产制造装备运维保养不当、缺乏有效的运行监管机制、人为破坏等原因，使工业 OT 域大量存在数据缺失、日志缺失和工业经验知识丢失等问题，造成生产制造装备运行数据的不可测和不可控，这是造成对工业现场运行状况误判的重要原因之一。此外，因控制器无法准确地获取运行数据，导致过程控制设备将错误的控制指令上传下达，则不仅会影响工业企业对运营状况的正确决策，而且会对生产经营埋下严重的安全隐患。

数据一致性差。对于来源于 IT 域的工业大数据，由于工业企业现有的信息化系统均不同程度的存在“系统林立”的问题，难以在数据生产过程中采用有效的控制手段来保障各个工业 IT 系统所产生的数据的一致性，当出现数据不一致时，数据使用者必须做出权衡，修整数据使其与原数据一致，或牺牲数据质量来保持与原数据一致性。

对历史数据缺乏“再生”机制。工业历史数据的“再生”将有助于提升工业全生命周期管理能力。传统的数据管理模式，历史数据往往在其生命周期的后期转为冷存储或损毁。而在工业大数据分析和应用中，实时数据与历史数据的应用具有各自的新价值。如何实现历史数据与实时数据的有效整合，实现工业大数据在工业过程、工业制成品全生命周期中的应用，将是数据质量管理的一个新的挑战。

缺乏标准化的数据质量管理框架。在不同的业务场景中，各种结构化和非结构化数据集被多个使用者共享和使用。然而，一方面，由于数据质量缺乏标准化管理的约束；另一方面，由于各类工业信息系统彼此割裂；同类型、不同格式的

数据在各类彼此独立的信息系统中运行，这意味着每种应用都有各自的数据使用方式，相同数据集在不同业务场景下的诠释不同，数据存在有效性和一致性方面的隐患。

数据质量补偿手段尚未得到广泛应用。目前，对于工业大数据面临的数据失真和失准问题，更多是基于经验进行修复，常用的数据滤波、平滑、降噪、插值及关联补偿等数据质量补偿手段尚未实现面向应用场景的固化；对于数据冗余、同类不同源数据不一致的问题，也缺乏包括统一数据 ID 管理及主数据管控等管理手段来促进多源异构数据的融合。过渡依赖经验的数据质量管理模式严重制约了智能制造的发展。

### 5.3.2 工业大数据质量检测方法

工业大数据的质量检测从包括筹备、采集、存储、共享、维护、应用、退役等各阶段的生命周期采取识别、衡量、监控和告警等管理措施，达到减少乃至杜绝工业数据的不完整、不准确、重复、不一致及复用性差等问题。工业大数据的质量检测基于不同的应用场景，围绕固有质量度量维度、环境质量度量维度、表达质量度量维度和可访问性质量度量维度，分别开展数据资产等级、数据性质、测点设计、监控点设计等数据质量检测工作，同步建立数据质量检查流程及持续改进机制。

### 5.3.3 工业大数据质量管理

工业大数据的质量管理需要工业企业建立完善的工业大数据质量管理组织架构，明确数据权属、管理者、使用者；面对不同的工业大数据质量问题，制定质量问题的定义、等级、处理及复盘机制，制定规范的数据质量改善流程，形成面向多样化的工业大数据应用场景的数据质量管理闭环。具体包括：

完善工业大数据质量管理组织架构。明确数据质量管理的重要性，分析其投资回报，克服组织障碍，建立完善的数据质量管理和问责机制是数据质量管理的必要条件。

建立工业大数据质量问题的响应机制。工业大数据质量管理属跨职能、跨领



域、跨平台的工作，任何一个环节出现问题都会影响数据质量。工业大数据质量问题的解决效率，直接影响数据的最终应用。需要针对数据质量问题定义不同的响应级别和要求，确保相关组织或团队能够第一时间采取措施，消除问题。

构建工业大数据质量改善流程。从数据质量需求定义、评估、分析、提升和监控五个流程环节持续改善工业大数据的数据质量。

应用工业大数据质量管理工具。选择合适的数据库质量管理工具支撑数据质量持续化的监控、记录、问题发现和优化。尽可能将流程规范系统化、质量衡量标准和结果系统化、质量问题系统化和质量问题追踪系统化。

## 5.4 主数据管理

目前，在业界存在着主数据不同角度的定义。IBM 公司在其有关主数据管理的红皮书《Master Data Management: Rapid Deployment Package for MDM》认为，所谓主数据是有客户、供应商、产品和账户的企业关键信息；有观点将主数据定义为“表示‘跟踪事物状态’的数据”；也有观点认为，企业主数据是用来描述企业核心业务实体的数据，是统一、稳定、共用的基础数据，比如客户、合作伙伴、员工、产品、物料单、账户等，它是具有高业务价值的、可以在企业内跨越各个业务部门被重复使用的数据，并且存在于多个异构的应用系统中。

基于上述定义，本白皮书对于主数据的定义进行了全面的概况：主数据是指满足跨部门业务协同需要的、反映核心业务实体状态属性的企业（组织机构）基础信息。

主数据主要包括配置型主数据和核心主数据。配置型主数据是描述业务或核心主数据属性分类的参考信息，会在整个组织内共享使用。一般依据国际标准、国家标准、行业标准或企业标准和相关规范等，在系统中一次性配置使用的基础数据，例如国家、民族、性别等规范性表述。配置型主数据相对稳定，不易变化。核心主数据指用来描述企业核心业务实体的主数据，是企业核心业务对象、交易业务的执行主体，如产品、物资、设备、组织机构、员工、供应商、客户、会计科目等。核心主数据是企业信息系统的神经中枢，是业务运行和决策分析的基础。核心主数据是相对“固定”的，变化缓慢。

主数据建设在企业信息化战略中处于核心地位，是企业信息化建设的基础支撑。一个企业信息化建设的水平与其主数据管理水平息息相关。主数据管理（MDM，Master Data Management）是一系列规则、应用和技术，用以协调和管理与企业的核心业务实体相关的系统记录数据。主数据管理的关键活动包括：理解主数据的整合需求，识别主数据的来源，定义和维护数据整合架构，实施主数据解决方案，定义和维护数据匹配规则，根据业务规则和数据质量标准对收集到的主数据进行加工清理，建立主数据创建、变更的流程审批机制，实现各个关联系统与主数据存储库数据同步，方便修改、监控、更新关联系统主数据变化。

主数据管理通过对主数据值进行控制，使得企业可以跨系统的使用和共享来自权威数据源的、协调一致的、高质量主数据，降低数据解析成本和复杂度，从而支撑跨部门、跨系统数据融合应用。

#### 5.4.1 主数据管控体系

针对企业主数据管理，首先需要建立有效地主数据管控体系。主数据管控体系主要由主数据管理制度、主数据管理组织、主数据管理流程、主数据管理评价等方面构成。

##### （1）主数据管理制度

主数据管理制度规定了主数据管理工作的内容、程序、章程及方法，是主数据管理人员的行为规范和准则。主数据管理制度一般包含《主数据管理办法》、《主数据标准规范》、《主数据提报指南》、《主数据维护细则》、《主数据管理系统操作手册》等。

##### （2）主数据管理组织

主数据管理组织主要包括企业内各类主数据的管理组织架构、运营模式、角色与职责规划，通过组织体系规划建立明确的主数据管理机构和组织体系，落实各级部门的职责和可持续的主数据管理组织与人员。

典型的主数据管控组织主要包含三层组织架构：决策层，设立主数据联合工作组，该工作组为非常设组织，其主要负责根据主数据管控流程、职责等，对主

数据相关重点问题进行联合决策；管理层，主要是信息化管理部门，负责主数据标准规范的制定工作，包括主数据规范的管理和发布，负责对主数据的日常管理；执行层，包括专职主数据管理员以及职能部门的兼职主数据管理员，其中专职主数据管理员负责主数据相关的系统改造、主数据清理、主数据稽核等具体工作，职能部门主数据管理员负责本部门主数据相关的协调、沟通、召集、以及主数据系统的日常操作。

### （3）主数据管理流程

主数据管理流程是提升主数据质量的重要保障，通过梳理数据维护及管理需求，建立符合企业实际应用的主数据管理流程，保证主数据标准规范得到有效执行，实现主数据的持续性长效管理。

主数据管理流程主要包含三方面内容：主数据业务管理，对主数据的申请、校验、审核、发布、变更、冻结、归档等进行全生命周期管理，满足主数据在企业深入应用的不同管理需求；主数据标准管理，通过对主数据标准的分析、制定、审核、发布、应用与反馈等流程进行设计，保证主数据标准的科学、有效、适用；主数据质量管理，对主数据的创建、变更、冻结、归档等业务过程进行质量管控，设计数据质量评价体系，实现数据质量的量化考核，保障主数据的安全可靠。

### （4）主数据管理评价

主数据管理评价是用来评估及考核主数据相关责任人职责的履行情况及数据管控标准和数据政策的落实情况，通过建立定性或定量的主数据管控评价考核指标，加强企业对主数据管控相关责任、标准与政策执行的掌控能力。

主数据管理评价指标从管控标准、数据认责和数据政策三个角度考虑，由数据所有人与数据认责人共同确定，通过定义一系列的衡量指标和规则，一方面落实和检查主数据的应用情况，另一方面考察和评估主数据管控、主数据标准、主数据质量的执行情况。

为了进一步保障主数据管理平台的成功实施和有效运行，必须做到组织、职能、责任、人员的四方面落实。制定涉及主数据管理的各个环节、组织、人员的一套绩效考核办法，明确各组织部门的职责与分工。

表 5-1 主数据管理评价考核指标

序号	考核方向	技术指标	衡量标准
1	及时性	及时率	满足时间要求的数据总数/总数据数
2	真实性和准确性	数据真实率	1-数据中失真记录总数/数据总记录数
		有效值比率	1-超出值域的异常值记录总数/总记录数
		流转过程失真率	数据传输失真记录总数/总记录数
		重复数据比率	重复记录数/总记录数
3	一致性	外键无对应主键的记录比率	外键无对应主键的记录总数/总记录数
		主数据一致率	一致的主数据总数/主数据总数
4	完整性	字段的空值率	空值记录总数/总记录数
		信息完备率	能够获取的指标数/总需求指标数

#### 5.4.2 主数据应用管理

基于主数据管控体系的建设，为保障实际工业数据管理需求的满足，还需加强对企业的主数据应用的管理。

(1) 明确管理要求。制定主数据应用管理制度规范，对主数据的应用范围、应用规则、管理要求和考核标准作出明确规定，并以此为依据，对主数据应用进行有效管理。

(2) 实施有效管理。主数据应用点多、线长、面广，管理难度很大，要实施有效管理，就必须要有健全的制度和可行的手段，在关键控制节点实施重点管控。一是加强宣讲和引导，通过业务主管部门落实好管理职责，要分工明确，责任到人，强化岗位责任制和考核管理，不能有管理死角。二是将信息系统建设项目实施主数据专项评审，确保信息系统在主数据应用方面符合管理要求。三是在客户管理、订单管理、结算管理等集中控制点实施主数据核验，对业务环节涉及的主数据进行全面核查，确保主数据在业务环节被有效使用；如有违规，进行必要的处罚。

(3) 强化服务保障。依靠便捷、可靠的主数据服务为主数据应用提供保障，包括主数据查询、主数据同步、主数据申请和主数据调用。有条件的单位可将主

数据服务深入到业务流程，从业务端发起请求，驱动主数据管理和数据服务，形成管理和应用的有机协同。

## 5.5 数据安全与隐私保护

工业大数据蕴涵着工业生产的详细情况及运行规律，也承载了大量市场、客户、供应链等信息，是工业互联网的核心要素，工业大数据安全管理因此成为工业互联网安全保障的重要任务之一。工业大数据安全管理的目的在于：一方面推动建立工业互联网全产业链数据安全管理体系，明确相关主体的数据安全保护责任和具体要求，加强数据生命周期各环节的安全防护能力，避免用户隐私或重要工业数据遭到不法窃取或利用；另一方面建立工业数据分级分类管理制度，形成工业互联网数据流动管理机制，明确数据留存、数据泄露通报要求；此外，还需通过加强监督检查落实企业的数据安全保护责任。

### 5.5.1 工业大数据安全问题

大数据技术应用于工业领域给企业带来巨大的效益，然而工业大数据对工业企业来说既是机遇也是挑战，在给企业带来巨大经济利益的同时，其本身所存在的安全问题也让企业面临着巨大的风险。一方面，由于工业控制系统的协议多采用明文形式、工业环境多采用通用操作系统且不及时更新、从业人员的网络安全意识不高，再加上工业数据的来源多样，具有不同的格式和标准，使其存在诸多可以被利用的漏洞。另一方面，在工业应用环境中，对数据安全有着更高的要求，任何信息安全事件的发生都有可能威胁工业生产运行安全、人员生命安全甚至国家安全等。因而，研究工业大数据安全管理，加强对工业企业的安全保护变得尤为重要。

### 5.5.2 工业大数据安全防护技术

工业大数据安全是跨多工业领域与学科的综合性问题，需要结合法律法规、行业特点、工业技术等多维度进行研究。考虑到工业大数据平台所承载的工业数据的巨大价值，因此这里将整个工业大数据安全技术体系分为工业大数据接入安全、工业大数据平台安全、工业大数据应用安全三个层次。其中工业大数据接入

安全为工业现场数据的采集、传输、转换流程提供安全保障机制；工业大数据平台安全为工业数据存储、计算提供安全保障基础；工业大数据应用安全为上层应用的接入、数据访问等提供强力的安全管控。

### （1）工业大数据接入安全

工业大数据接入安全必须保障工业边缘设备实时数据采集、工业远程状态监控、工业企业系统数据抽取等从外部系统获取工业数据，并进行清洗、转换、传输以进入工业大数据平台的完整数据传输链的安全。

数据采集端支持采集模块的注册及安全认证机制，保障数据采集应用的合规性，采集数据的准确性；边缘计算模块支持统一模块管理下发及签名校验机制，保障数据预处理应用的合法性和可靠性；数据传输通路支持通道加密，保障传输过程中的机密性和完整性。

### （2）工业大数据平台安全

工业大数据平台安全是对工业数据资源的存储、访问、运算等功能的安全保障，包括数据的存储安全、计算安全、平台管理安全以及基础设施安全。

平台存储安全支持数据多备份设置与恢复机制，并采用数据访问控制机制来防止数据的越权访问；计算安全支持计算发起方的身份认证和访问控制机制，确保只有合法的用户或应用程序才能发起数据处理请求；平台管理安全包括平台组件的安全配置、资源安全调度、补丁管理、安全审计等，确保整个平台组件及运行状态安全可控，同时还应强化平台的数据隔离和访问公职机制，实现数据“可用不可见”；平台软硬件基础设施安全包括基础网络安全、虚拟化安全等，从而保障整个大数据平台的安全运行。

### （3）工业大数据应用安全技术

工业大数据应用会对存储于工业大数据平台的海量数据进行查询、分析、计算、导出等操作，因此在工业大数据平台提供数据服务的同时，其安全风险也随之被暴露，攻击者可利用各类已知或未知漏洞发起攻击，达到破坏系统或者获取数据信息的目的，因此需要对数据应用安全进行严格管控。

工业大数据应用安全应从几方面考虑：支持应用访问签名机制，确保只有授权的应用才能提交数据访问请求；支持应用数据按需访问，避免数据访问范围的扩大化；支持应用行为实时监控，实时拦截应用中包含的攻击行为，包括数据访问范围、频率、SQL 语句合法性等；建立完整的应用流程管理机制，包括应用的提交、执行、状态监控、结果审计等，确保每个应用的审批、控制与追责有效结合，避免高权限人员的恶意操纵或误操作行为；构建完备的应用测试环境及测试规范，确保只有符合安全策略的应用可以审批执行。

### 5.5.3 工业大数据安全管理机制

针对工业大数据安全问题，在技术范畴外还应通过管理机制实现工业大数据安全防护的具体落实。从管理角度看，工业大数据安全管理机制的构建需要考虑以下方面：

(1) 工业大数据安全管理策略。以政府、企业、机构、专家团队等利益相关方共同参与，建立健全的数据分级分类安全治理结构，明确工业大数据安全管理的指导思想、基本原则、重点任务、保障支撑等关键环节的工作内容、沟通机制和工作规范。

(2) 工业大数据分级分类管理。针对工业数据的来源、内在价值、内容敏感程度、访问范围等差异，使得数据的管理，特别是安全策略异常复杂。因此对数据进行分级分类，从而实现更精细化的数据安全治理，避免安全策略上的“一刀切”，是实现工业大数据安全治理和使用的有效方法。

(3) 工业大数据安全管理组织机构。工业企业需要建立工业大数据安全管理组织，并由该组织建立大数据安全管理保障制度和规范，推动安全策略的统一实施，做到安全策略集中管理、集中修订、集中更新。安全管理组织还将建立对外合作安全管理(外部业务合作和外部代维代建等)、内部安全管理、数据分类分级管理、应急响应机制、资产设施保护和认证授权管理等相关制度。

(4) 数据安全审计与稽核。数据安全稽核是安全管理部门的重要职责，以此保障数据安全治理的策略和规范被有效执行和落地，以确保快速发现潜在的风险和行为。具体包括：行为审计与分析，建设数据安全防护体系时必须具备审计

能力，利用数据库协议分析技术将所有访问和使用数据的行为全部记录下来，包括账号、时间、互联网协议（IP）地址、会话、操作、对象、耗时、结果等内容；权限变化监控，权限变化监控是指监控所有账号权限的变化情况，包括账号的增加和减少，权限的提高和降低，这对抵御外部提权攻击，对抵御内部人员私自调整账号权限进行违规操作均是必不可少的关键能力；异常行为分析，在安全稽核过程中，除了明显的攻击行为和违规的数据访问行为外，很多的数据入侵和非法访问是掩盖在合理的授权下的，这就需要通过一些数据分析技术，对异常性的行为进行发现和定义。



## 6 工业大数据的标准体系

### 6.1 工业大数据标准化工作基础

#### 6.1.1 国际标准化

目前大数据国际标准化工作主要集中在 ISO/IEC JTC 1 进行。ISO/IEC JTC 1/SC 32 数据管理和交换分技术委员会，主要致力于研制信息系统环境内及之间的数据管理和交换标准，为跨行业领域协调数据管理能力提供技术性支持，其标准化技术内容涵盖：协调现有和新生数据标准化领域的参考模型和框架；负责数据域定义、数据类型和数据结构以及相关的语义等标准；负责用于持久存储、并发访问、并发更新和交换数据的语言、服务和协议等标准；负责用于构造、组织和注册元数据及共享和互操作相关的其他信息资源（电子商务等）的方法、语言服务和协议等标准。ISO/IEC JTC 1/SC 42/WG 2（原 ISO/IEC JTC 1/WG 9）大数据工作组，主要致力于大数据领域的国际标准化研究，具体包括开展大数据领域术语与定义、关键技术、参考模型以及用例等基础标准研究，评估分析当前大数据标准的具体需求，研制大数据标准化路线图等。

工业大数据的国际标准化工作，目前仍集中于传统的工业控制领域所涉及的工业数据的标准化工作。如 ISO/TC 184 自动化与集成委员会下设 SC 4 工业数据分委会，围绕工业自动化系统和集成中的产品数据表示和交换、特性数据交换、工业制造管理数据以及数据质量等方面开展了国际标准研制。

#### 6.1.2 国内标准化

我国工业大数据标准化工作主要依托全国信标委大数据标准工作组工业大数据专题组开展。工业大数据专题组于 2015 年 7 月成立，是全国信标委大数据标准工作组（SAC/TC 28/WG 22）下属专题组之一，主要围绕大数据技术在工业产品的研发、生产、服务等方面的应用以及工业数据管理等方面开展标准化研究，致力于以标准化的手段推动我国制造业向网络化、智能化、绿色化、服务型升级。

前期，工业大数据专题组广泛征集标准化需求，持续推进工业大数据国家标

准研制。目前《信息技术 工业大数据 术语》（20180988-T-469）、《信息技术 大数据 工业应用参考架构》（20173819-T-469）、《信息技术 大数据 产品要素基本要求》（20173820-T-469）、《智能制造 工业数据空间模型》（20182054-T-339）、《智能制造 多模态数据融合系统技术要求》（20182040-T-339）、《智能制造 工业大数据平台通用要求》（20182053-T-339）、《智能制造 工业大数据时间序列数据采集和存储框架》（20182052-T-339）等 7 项国家标准已完成立项，部分标准已经在工业大数据产业应用联盟等相关单位开展了试点示范和应用推广工作。

2015 年 12 月，工业和信息化部、国家标准化管理委员会联合发布了《国家智能制造标准体系建设指南（2015 年版）》（以下简称“智能制造建设指南”）并建立动态更新机制，目前“建设指南”已更新至 2018 版。“智能制造建设指南”明确提出了建设智能制造标准体系的总体要求、建设思路、建设内容和组织实施方式，给出了智能制造标准体系框架，框架中明确了工业大数据属于智能制造标准体系中五大智能赋能技术之一。“智能制造建设指南”中对于工业大数据标准给出了相关描述，主要包括平台建设的要求、运维和检测评估等工业大数据平台标准；工业大数据采集、预处理、分析、可视化和访问等数据全生命周期技术标准；数据质量管理、数据资产管理、数据管理能力评估等数据管理标准；企业内部数据流通、企业外部数据共享交换等数据互联互通标准。主要用于典型智能制造模式中，提高产品全生命周期各个环节所产生的各类数据的处理和应用水平。而其中对应的具体标准研制任务主要由全国信标委大数据标准工作组工业大数据专题组组织承担。

2019 年 1 月，工业和信息化部、国家标准化管理委员会联合发布了《工业互联网综合标准化体系建设指南》（以下简称“工业互联网指南”）。“工业互联网指南”梳理了当前工业互联网技术发展现状，提出了工业互联网标准体系的建设思路、建设目标、建设内容以及组织实施方式，给出了工业互联网标准体系框架。“工业互联网指南”中将工业大数据归纳于“总体标准”中“平台与数据标准”部分，并梳理了当前工业大数据相关国家标准共 8 项，其中已发布国家标准 4 项，在研国家标准 4 项，相关标准均由全国信标委大数据标准工作组负责归口研制。

此外，全国自动化系统与集成标准化技术委员会工业数据分技术委员会（SAC/TC 159 SC 4）也开展一些与产品生产相关的底层数据标准化工作，其专业范围是制造应用数据和语言方面的标准化工作，包括产品数据表达和交换、零件库结构、制造管理数据和系统集成等，但不能满足智能制造环境下的工业大数据标准化需求。

## 6.2 工业大数据标准体系框架

基于我国工业大数据技术、产业发展现状，以及已有工业大数据标准化基础，结合大数据在工业领域应用特点、典型应用场景以及未来发展趋势，本白皮书对《工业大数据白皮书（2017版）》中提出的标准体系进行修订，形成工业大数据标准体系框架，如图6-1所示。

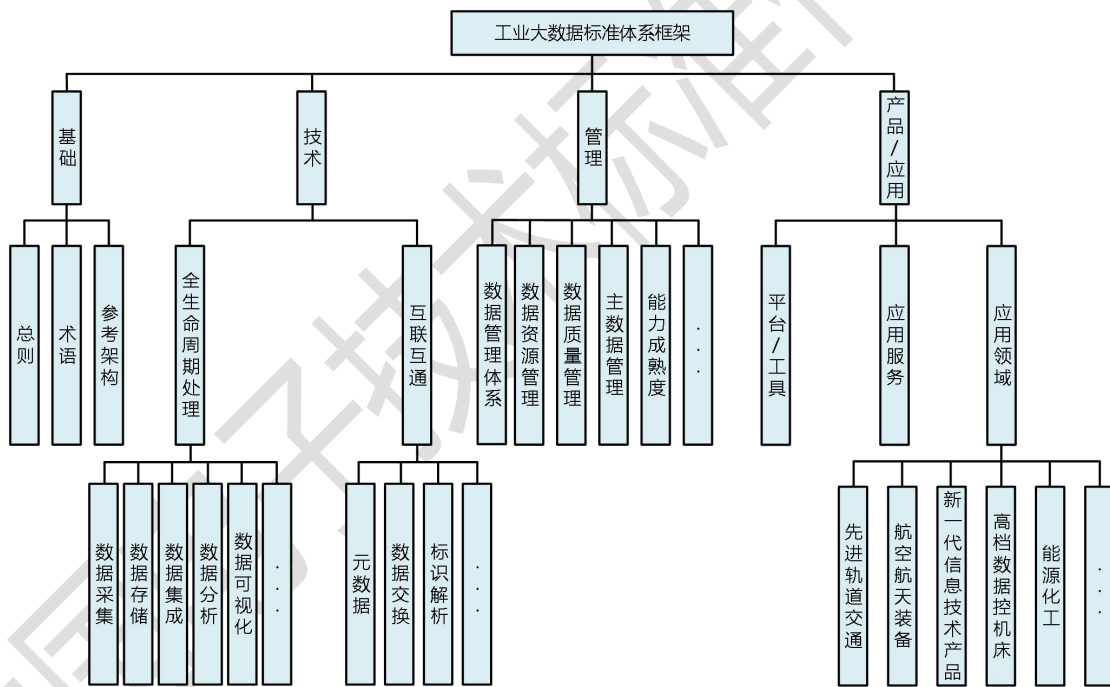


图 6-1 工业大数据标准体系框架

工业大数据标准体系主要包括基础标准、技术标准、管理标准、产品/应用标准等四个部分。

### 1) 基础标准

基础标准主要用于统一工业大数据相关概念，解决工业大数据基础共性关键问题，包括总则、术语和参考模型等。术语用于统一和规范工业大数据领域的的

通用术语及其定义，参考模型则规范了工业大数据的基础架构和研究范围。

## 2) 技术标准

技术标准主要用于对工业大数据关键技术进行规范，包括全生命周期处理技术、互联互通技术两部分。工业数据全生命周期处理技术标准主要针对工业大数据在某一应用场景内从产生到其使用、终止（回收）这一过程的关键技术进行标准化，包括数据采集、数据存储、数据集成、数据分析、数据可视化等标准；互联互通技术标准主要针对工业大数据在各个业务系统间以及不同平台间流通所涉及到的技术要求进行规范，包括元数据标准、数据交换标准、标识解析标准等。

## 3) 管理标准

管理标准主要对工业大数据在应用过程中的数据管理方法、流程、机制进行规范，包括工业大数据数据管理体系、数据资源管理、数据质量管理、主数据管理、能力成熟度等。其中数据管理体系主要针对工业数据管理的参与角色、管理架构、管理办法等进行规范。数据质量标准主要针对工业数据质量制定相应的指标要求和规格参数，确保工业数据在产生、存储、交换和使用等各个环节中的质量水平，包括定义业务需求及相关业务规则，数据质量检测、质量评价、数据溯源等标准。数据资源管理主要针对工业大数据在实际应用中的数据资源分类、数据资源目录建设、数据资源优化等方面进行规范。主数据管理主要针对工业企业在业务应用中涉及到的主数据管理的相关技术、流程进行标准化规范。能力成熟度标准主要对工业企业数据管理能力的建设、评估所涉及到的技术手段和流程进行规范，支撑企业开展数据管理能力自评和第三方评测。

## 4) 产品/应用标准

产品/应用标准主要针对工业大数据系统产品、工业大数据应用服务以及工业大数据在垂直领域应用中涉及的技术进行标准化规范，包括平台/工具、应用服务、应用领域三个部分。其中平台/工具标准主要针对工业大数据存储、处理、分析、融合等系统或工具从技术架构、功能要求、性能规范、平台接口、管理维护等方面进行规范；应用服务标准主要针对工业企业对外提供的工业数据解决方案进行评估方法、评估流程进行规范；应用领域标准则主要针对工业大数据在垂直领域应用中根据其领域或典型应用场景特性产生的专用数据标准。

### 6.3 工业大数据标准明细表

根据当前工业大数据国家标准现状，现整理工业大数据标准明细表如下：

表 6-1 工业大数据标准明细表

序号	一级分类	二级分类	标准/计划号	标准名称	状态
1.	基础标准	总则	——	信息技术 工业大数据 标准化指南	拟研制
2.		术语	20180988-T-4 69	信息技术 工业大数据 术语	在研
3.		参考架构	20173819-T-4 69	信息技术 大数据 工业应用参考架构	在研
4.	技术标准	全生命周期处理	——	信息技术 工业大数据 装备数据采集规范	拟研制
5.			20182052-T-3 39	智能制造 工业大数据时间序列数据采集和存储框架	在研
6.		互联互通	——	信息技术 工业大数据 订单元数据	拟研制
7.			20173820-T-4 69	信息技术 大数据 产品要素基本要求	在研
8.			——	工业大数据 工业信息交换规范	拟研制
9.			20170057-T-4 69	智能制造 对象标识要求	报批
10.			20182054-T-3 39	智能制造 工业数据空间模型	拟研制
11.	管理标准	数据质量管理	——	信息技术 工业大数据 数据质量评价指南	拟研制
12.		主数据管理	——	信息技术 工业大数据 主数据管理规范	拟研制
13.		能力成熟度	——	工业数据管理成熟度模型	拟研制
14.	产品/应用标准	工具/平台	20182053-T-3 39	智能制造 工业大数据平台通用要求	在研
15.			20182040-T-3 39	智能制造 多模态数据融合系统技术要求	在研

在工业大数据领域，目前已开展 8 项国家标准研制，涉及术语、参考架构、全生命周期处理、互联互通、工具/平台等部分。下一步拟重点围绕工业大数据管理领域推进国家标准研制工作，开展主数据管理、数据质量管理、工业数据管理成熟度模型等国家标准研制，以标准为抓手，为企业提供数据管理相关体系架构和技术支撑，帮助企业构建数据管理及其相关解决方案能力建设，不断完善工

业企业的数据管理能力，提升行业数据管理水平。

## 6.4 工业大数据重点标准描述

本节将对我国工业大数据领域重点国家标准进行介绍：

(1) 《信息技术 大数据 工业应用参考架构》国家标准主要规范了大数据在工业应用中的适用范围、参考架构。具体地，将围绕工业大数据典型应用场景和产品全生命周期各环节，确定工业大数据领域的数据架构、技术架构和平台参考架构。本标准适用于未来工业大数据的规模应用参考和依据。

(2) 《智能制造 工业大数据平台通用要求》国家标准对工业大数据平台建设和应用中涉及到的术语、定义和概念进行解释，提出工业大数据平台建设和应用对企业系统、资源和能力的基本要求，以及实现工业大数据平台建设和应用的前提条件。具体对工业大数据平台的通用模型、功能构建等内容进行规范。本标准适用于工业大数据平台的设计、开发、测试、验收和应用。

(3) 《智能制造 工业数据空间模型》国家标准主要针对工业数据交换共享融合的基础和共性问题，引入了“工业数据空间”的概念，致力于建立应用在工业领域的，受信任、可共享的数据空间，以达到促进数据提供者和数据使用者之间实现安全、可信的数据交换的目的。本标准从商业、数据、服务、软件、安全五个方面规范了工业数据空间总体通用功能的建设要求和建设内容。本标准适用于为工业数据空间建设提供总体的规划架构思路和共性基础功能需求。

(4) 《智能制造 多模态数据融合系统技术要求》国家标准主要针对当前智能制造产业中对大量多模态数据进行自动融合分析的迫切需求，面向业内不同厂商的多模态数据融合系统的基本架构，功能要求以及其他非功能性要求进行统一规范。本标准适用于指导面向智能制造领域的多模态数据融合系统产品的研制、开发、测试、评估和采购。

## 7 下一步工作建议

### 7.1 强化关键技术，提升工业大数据平台能力建设

开展工业大数据关键技术研究，深化工业大数据采集、存储、处理、分析等核心技术创新。结合各工业领域典型需求，突破工业大数据机理模型建模技术、知识推理技术、仿真测试、并行分析处理等高端新型工业软件核心技术，实现对海量工业数据的可靠存储与处理，提高国产高端工业大数据产品和系统供给能力，并有效保障工业数据与工业系统的安全性。加强工业大数据平台能力建设，构建安全、可信的工业大数据平台联通机制，建立行业认可的工业大数据平台和评估机制；选择重点领域、重点行业，依托工业大数据平台，打通全产业链数据壁垒，实现数据驱动的工业制造企业商业模式创新。

### 7.2 加强工业大数据管理体系建设，提升数据资源价值

推动建设“国家层面、行业层面、企业层面”自上而下、多元共治的工业大数据管理体系。国家层面，加强顶层设计，研究数据管理政策指引，完善数据确权、数据资产等相关制度及法律法规，加强工业数据管理能力成熟度评估，建立完善国家级工业大数据管理体系；行业层面，梳理工业数据资源目录清单，打破数据壁垒，完善产业链上下游数据交换共享机制，促进工业数据流通和利用；企业层面，将工业大数据管理纳入企业管理范畴，构建企业数据管理机制，制定和实施系统化的工业数据管理制度、流程和方法，推进企业内部数据资源整合，建立企业数据管理能力自查评测机制，进而提高全行业数据管理水平。

### 7.3 持续完善工业大数据标准体系，推动标准落地实施

加强工业大数据标准化顶层设计，建立并不断完善工业大数据标准体系，发挥标准化对工业数字经济的重要支撑作用。针对数据集成、互联互通等关键瓶颈问题，重点开展工业数据管理、数据交换共享以及工业大数据平台等基础标准研制。完善标准应用环境，选择重点行业和地方对重点标准先行先试，在工业数字经济中率先探索工业大数据标准应用模式，加强标准的试点示范和应用推广。建立工业数据管理标准符合性评估体系，强化标准对市场培育、服务能力提升和行

业管理的支撑作用。

#### 7.4 探索工业大数据创新应用示范，构建产业生态体系

遵循以“需求为导向，应用促发展”的工作思路，凝聚业内优秀产、学、研、用资源，选择基础条件好、示范效应强、影响范围广的行业领域积极开展创新应用示范，构建多位一体的、动态的产业生态体系，进一步探索工业大数据产业的新模式、新业态，打造可复制推广的实践样板。随着智能化改造的推进，逐步扩大产业生态范围，打破原有产业界限，不断扩大业内影响范围，最终建立形成覆盖全产业链的生态模式，为最终实现我国数字经济建设提供管理和实践经验。