

UDC

MH

中华人民共和国行业标准

P

MH/T 5067—2023

智慧民航数据治理规范 数据治理技术

Specifications for smart civil aviation data governance
—Data governance technology

2023-04-25 发布

2023-06-01 施行

中国民用航空局 发布

中华人民共和国行业标准

智慧民航数据治理规范 数据治理技术

Specifications for smart civil aviation data governance
—Data governance technology

MH/T 5067—2023

主编单位：中国民用航空局发展计划司

中国民航管理干部学院

中国民用航空局信息中心

批准部门：中国民用航空局

施行日期：2023年6月1日

中国民航出版社有限公司

2023 北 京

图书在版编目 (CIP) 数据

智慧民航数据治理规范. 数据治理技术 / 中国民用航空局发展计划司, 中国民航管理干部学院, 中国民用航空局信息中心主编. —北京: 中国民航出版社有限公司, 2023. 4

ISBN 978-7-5128-1214-7

I. ①智… II. ①中… ②中… ③中… III. ①民用航空-数据管理-管理规范-中国 IV. ①F562-65

中国国家版本馆 CIP 数据核字 (2023) 第 063990 号

中华人民共和国行业标准
智慧民航数据治理规范 数据治理技术
MH/T 5067—2023

中国民用航空局发展计划司
中国民航管理干部学院 主编
中国民用航空局信息中心

责任编辑 韩景峰

出版 中国民航出版社有限公司 (010) 64279457
地址 北京市朝阳区光熙门北里甲 31 号楼 (100028)
排版 中国民航出版社有限公司录排室
印刷 北京金吉士印刷有限责任公司
发行 中国民航出版社有限公司 (010) 64297307 64290477
开本 880×1230 1/16
印张 2
字数 55 千字
版印次 2023 年 5 月第 1 版 2023 年 5 月第 1 次印刷

书号 ISBN 978-7-5128-1214-7

定价 28.00 元

官方微博 <http://weibo.com/phcaac>

淘宝网店 <https://shop142257812.taobao.com>

电子邮箱 phcaac@sina.com

中国民用航空局 公告

2023 年第 8 号

中国民用航空局关于发布《智慧民航数据治理规范 数据治理技术》的公告

现发布《智慧民航数据治理规范 数据治理技术》（MH/T 5067—2023），自 2023 年 6 月 1 日起施行。

本标准由中国民用航空局发展计划司负责管理和解释，由中国民航出版社出版发行。

中国民用航空局

2023 年 4 月 25 日

前 言

数据作为新的生产要素，是民航行业数字化、网络化、智能化的基础。为充分发挥数据价值，完善行业数据治理体系，提升行业数据治理能力，民航局提出了行业数据治理顶层框架，确定了“7+1”智慧民航数据治理规范体系（即7部行业标准和1部信息通告），分别为民航数据治理《框架与管理机制》《数据架构》《数据安全》《数据质量》《数据服务》《数据共享》《数据治理技术》和典型实践案例。数据治理技术作为民航数据治理体系中的技术保障，决定着数据管理、处理与分析的能力与效率。为规范民航数据治理相关信息技术工具或功能的设计，指导民航各单位数据治理技术能力建设，制定本规范。

在本规范的编制过程中，编写组深入调研和总结民航行业数据治理的技术需求，借鉴行业内外的数据治理技术实践成果，经广泛征集行业意见和多次专家论证审查，最终形成本规范。

本规范共5章。主要内容包括：总则、术语和缩略语、民航数据治理技术一般要求、民航数据管理类技术要求、民航数据操作类技术要求。

本规范的日常维护工作由中国民航管理干部学院大数据与人工智能系负责，执行过程中如有意见或建议，请函告本规范日常维护组（地址：北京市朝阳区花家地东路3号；电话：010-58250678；邮箱：bigdata@camic.cn），以及民航工程建设标准化技术委员会秘书处（地址：北京市朝阳区惠新东街甲2号住总地产大厦；电话：010-64922342；邮箱：mhgcjsbwh@163.com），以便修订时参考。

主编单位：中国民用航空局发展计划司

中国民航管理干部学院

中国民用航空局信息中心

主 编：包 毅 刘 一

参编人员：袁 婷 韩兴军 李 波 朱国辉 刘心桥 程 宇

主 审：韩 钧 殷时军 于 剑

参审人员：高利佳 黄文强 杨建伟 王 欣 马 力 王瀚林 熊 朝

张 轶 闫 超 吴国华 李 冰 董战鲲

目次

1	总则	1
2	术语和缩略语	2
2.1	术语	2
2.2	缩略语	3
3	民航数据治理技术一般要求	4
4	民航数据管理类技术要求	5
4.1	数据标准管理技术要求	5
4.2	元数据管理技术要求	6
4.3	主数据管理技术要求	6
4.4	数据资产管理技术要求	7
4.5	数据质量管理技术要求	8
4.6	数据安全管管理技术要求	9
4.7	数据共享管理技术要求	10
5	民航数据操作类技术要求	12
5.1	数据采集技术要求	12
5.2	数据存储技术要求	13
5.3	数据计算技术要求	14
5.4	数据融合技术要求	14
5.5	数据分析技术要求	15
	标准用词说明	17
	引用标准名录	18

1 总 则

1.0.1 为提高民航数据治理效能，指导各单位数据治理技术能力建设，制定本规范。

1.0.2 本规范适用于民航数据治理技术能力建设工作的。

【条文说明】各单位在本规范内容的框架与指导下，结合自身发展现状及目标，进一步细化具体实施方案与细则。

1.0.3 民航行业数据治理技术除应满足本规范的规定外，尚应符合国家、行业现行有关标准的规定。

2 术语和缩略语

2.1 术 语

2.1.1 数据治理 data governance

数据治理是指数据资源及其应用过程中的相关管控活动，包括对数据进行处置、格式化和规范化的过程集合。

[MH/T 5054, 2.0.1]

2.1.2 数据标准 data standard

数据标准是指共同遵守的对数据含义和业务规则的规范性约束。

[MH/T 5055, 2.0.4]

2.1.3 元数据 metadata

元数据是指描述数据的数据，是关于数据的组织、数据域及其关系的信息。

[MH/T 5055, 2.0.7]

2.1.4 主数据 master data

主数据是指反映民航核心业务实体状态属性的基础数据，在数据价值链上应用于多个业务流程、部门和信息系统。

[MH/T 5055, 2.0.8]

2.1.5 数据资产 data asset

数据资产是指能够带来经济和社会效益的，以物理或电子的方式记录的数据资源，包括文件、资料、图表、电子数据等。

[MH/T 5055, 2.0.2]

2.1.6 结构化数据 structured data

结构化数据是指存储在数据库里，可以用二维表结构表示的数据。

[GB/T 37721, 3.2]

2.1.7 非结构化数据 unstructured data

非结构化数据是指除了结构化数据之外的没有明确结构约束的数据。

[GB/T 37721, 3.3]

2.1.8 批处理 batch processing

批处理是指将一个大型作业分解成为多个任务交由多个节点分别处理，再将分解后多个任务处理的结果汇总起来，得出最终的分析结果的计算框架，具备高可用、高扩展、高并发等能力。

[GB/T 37722, 3.2]

2.1.9 流处理 stream processing

流处理是指对具有实时、高速、无边界、瞬时性等特性的流式数据进行实时处理的计算。

[GB/T 37722, 3.3]

2.1.10 图处理 graph processing

图处理是指以“图论”为基础的对数据的一种“图”结构的抽象表达，以及在这种数据结构上的计算模式。

[GB/T 37722, 3.4]

2.2 缩略语

2.2.1 API (Application Programming Interface) 应用程序接口

2.2.2 SQL (Structured Query Language) 结构化查询语言

2.2.3 NoSQL (Not Only Sql) 非关系型数据库

2.2.4 OLAP (On-Line Analytical Processing) 联机分析处理

2.2.5 MPP (Massively Parallel Processing) 大规模并行处理

3 民航数据治理技术一般要求

3.0.1 民航数据治理技术包含数据管理类技术和数据操作类技术，其框架如图 3.0.1 所示。数据管理类技术是支持数据治理中数据相关管理活动的信息技术工具或功能。数据操作类技术是支持数据治理全流程过程中对数据进行操作的信息技术工具或功能。

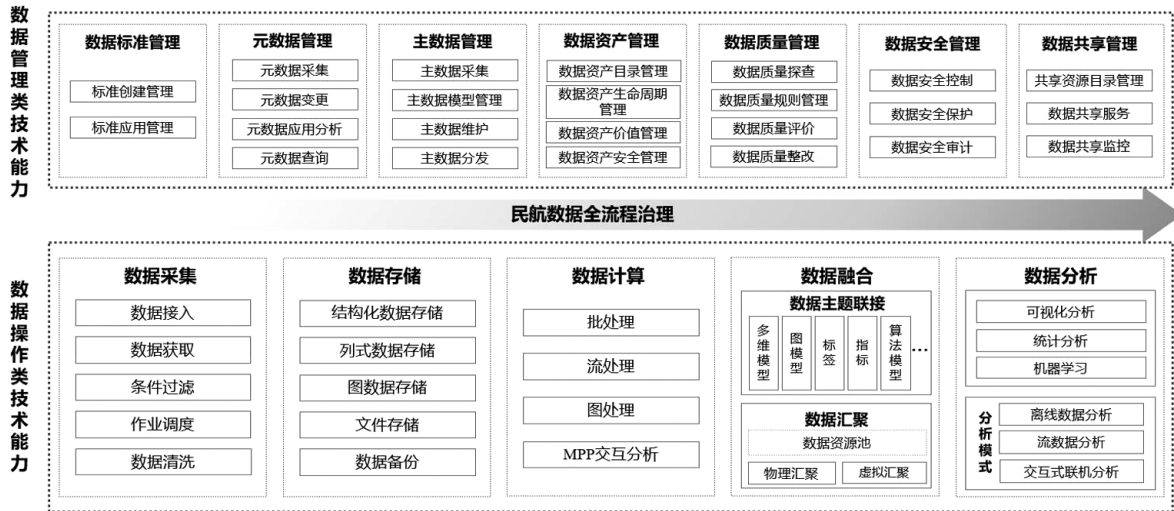


图 3.0.1 民航数据治理技术框架

3.0.2 民航数据管理类技术包括但不限于数据标准管理技术、元数据管理技术、主数据管理技术、数据资产管理技术、数据质量管理技术、数据安全治理技术、数据共享管理技术。

3.0.3 民航数据操作类技术包括但不限于数据采集技术、数据存储技术、数据计算技术、数据融合技术、数据分析技术。

3.0.4 民航数据治理技术应符合下列原则：

- 1 契合本单位的数据治理目标、范围及所治理数据的特征，并贯穿数据治理实施流程；
- 2 考虑本单位实际的信息系统建设和运营能力；
- 3 各类数据治理软件工具等应具备互联互通能力。

【条文说明】本规范“单位”指民航行业各级行政主体、企业、直属单位和行业协会等组织。

4 民航数据管理类技术要求

4.1 数据标准管理技术要求

4.1.1 数据标准管理技术用于支持数据标准的数字化创建和管理，规范数据资源的技术属性、业务属性和管理属性，保持数据口径的一致性。宜包括标准创建管理、标准应用管理等技术能力。

【条文说明】数据标准涉及的数据技术属性、业务属性、管理属性符合《智慧民航数据治理 数据架构》（MH/T 5055）中 5.1.1 的规定。

4.1.2 标准创建管理宜包括标准生成、标准映射、标准审核、标准维护、版本管理等能力，具体要求如下：

1 标准生成应支持按照领域、业务域、业务对象、逻辑数据实体、属性等生成标准细则，对于基础数据标准应支持数据的技术属性、业务属性、管理属性定义，对于指标数据标准应具备维度定义、度量定义等能力；

【条文说明】基础数据标准和指标数据标准符合《智慧民航数据治理 数据架构》（MH/T 5055）中 4.2.1、5.1.2 的规定。

2 标准映射应支持将制定的标准与实际数据进行关联映射，维护标准与元数据之间的落地映射关系，包括元数据与数据标准的映射、元数据与数据质量的映射，以及数据标准和数据质量的映射，应提供灵活的映射配置功能，并能对映射结果做页面展示；

3 标准维护应支持对标准状态进行管理，包括增删改、审核、定版、发布、废止等；

4 版本管理应支持对发布状态的标准进行版本管理，支持对已发布或已废止的标准进行变更轨迹查询。

4.1.3 标准应用管理宜包括标准检索、标准检查、标准建模、标准分析等能力，具体要求如下：

1 标准检索应实现对本单位所有标准按编号、名称、描述等进行检索；

2 标准检查应支持对元数据进行标准符合性检测，并分析标准符合性检测通过率；

3 标准建模应支持标准与对应表的映射，实现基于标准的对应表模型设计；

4 标准分析应支持对标准的分布情况、使用情况、访问情况等分析，为标准的修订提供参考依据。

4.2 元数据管理技术要求

4.2.1 元数据管理技术通过元数据自动化采集,支持绘制数据地图、标明数据分布、分析数据关系、管理模型变更。宜包括元数据采集、元数据变更、元数据应用分析、元数据查询等技术能力。

4.2.2 元数据采集应具备以下能力:

- 1 提供多种数据源的元数据采集适配器,支持自动采集和手动采集两种模式;
- 2 支持对元数据采集任务、采集器状态、数据源参数等进行配置和维护,跟踪采集任务状态和查询采集执行日志等。

4.2.3 元数据变更应具备以下能力:

- 1 支持元数据变更监控,记录元数据的历史变更轨迹;
- 2 实现对变更的协同处理,当上游系统的数据模型发生变更时,对下游系统提供变更提醒及相关元数据信息;
- 3 对元数据的关键变更进行版本发布,当元数据发生变更后,可备份当前元数据后发布新版本。

4.2.4 元数据应用分析应具备以下能力:

- 1 支持数据地图绘制、血缘分析、影响分析和全链分析;
- 2 支持对某项元数据的上下游数据信息及其关系的可视化呈现。

4.2.5 元数据查询应支持按数据源等检索条件进行查询,快速定位数据。

4.3 主数据管理技术要求

4.3.1 主数据管理技术用于主数据的数字化定义、管理和共享,保障主数据在各个信息系统间的准确性、一致性和完整性。宜包括主数据采集、主数据模型管理、主数据维护、主数据分发等技术能力。

4.3.2 主数据采集应支持主数据统一、规范的采集和整合,具备以下能力:

- 1 应支持主数据识别、自动采集,通过 API 接口等方式从主数据源头业务系统自动采集主数据;

【条文说明】从源头业务系统自动采集主数据符合《智慧民航数据治理 数据架构》(MH/T 5055)中的规定。

- 2 应支持在主数据自动采集或手工录入过程中进行审批、质检；
- 3 应支持数据整合，依据主数据标准和主数据模型定义的规则进行跨部门、跨系统的主数据校验、清洗、发布，形成整合统一的、可信任的主数据；
- 4 宜支持主数据共享管理视图，能够直观、清晰地展示相应的主数据标准层级、数据标准对象、版本、主管部门、数据质量水平、涉及的系统列表、已对接和未对接主数据等相关信息。

4.3.3 主数据模型管理应具备以下能力：

- 1 以主数据标准为基准，支持通过可视化建模工具，定义主数据对象、编码规则、属性值和控制流程等基础要素，构建主数据标准模型；
- 2 提供编码定义、码段定义、编码维护等能力，根据编码要求自动生成主数据编码。

4.3.4 主数据维护应支持主数据创建、审批、发布、修改、失效等全生命周期的维护，应符合下列要求：

- 1 支持通过申请操作获取主数据；
- 2 支持相关维护人员的主数据修改权限审批，及其对主数据的修改；
- 3 支持对主数据质量的周期性检测及问题审计；
- 4 支持对主数据进行冻结、解冻和失效操作，冻结后的主数据不能更新及分发，解冻需要批准，解冻后的主数据能够再次使用，失效后的主数据不能再次使用。

【条文说明】冻结、解冻是常用的主数据变更操作。冻结的数据不可用，不能修改、移动和查看，解冻数据指让冻结的数据恢复正常。

4.3.5 主数据分发用于实现业务系统对主数据的需求，应具备以下能力：

- 1 支持批量的主数据分发服务，宜支持对接业务系统进行主数据批量分发；
- 2 支持主数据分发管理，提供查询接口、申请接口、修改接口、冻结接口、失效接口等。

4.4 数据资产管理技术要求

4.4.1 数据资产管理技术用于以资产价值视角提供数据管理功能。宜包括数据资产目录管理、数据资产生命周期管理、数据资产价值管理、数据资产安全管理等技术能力。

4.4.2 数据资产目录管理应具备以下能力：

- 1 支持对已识别的数据资产进行统一登记、统一分类、统一展示和查询；
- 2 实现多维度、多样式展示，例如数据资产树、数据资产地图、数据资产网络等；
- 3 支持对数据资产基本属性、业务要素、管理要素、价值要素等信息的登记。

【条文说明】基本属性包括数据的来源、类型、结构、规模、更新周期等；业务要素包括业务描述、业务指标、业务规则等；管理要素包括数据权属、分类分级、安全信息、数据溯源、职责

权限和应用情况等；价值要素包括领域信息、应用价值等。

4.4.3 数据资产生命周期管理应具备以下能力：

- 1 支持数据资产识别，发现和标识数据资产，并登记到数据资产目录；
- 2 支持数据资产盘点，确保数据资产对象与数据资产登记信息的一致性、完整性和准确性等；
- 3 支持数据资产变更，当数据资产管理活动或业务需求触发数据资产变化时，及时更新数据资产目录；
- 4 支持数据资产处置，通过数据资产的销毁、转移等，优化数据资产配置等数据资产处置手段，降低管理成本；
- 5 支持数据资产应用，围绕具体业务场景，在确保安全、合规的前提下，识别数据应用的途径和渠道，对数据资产内部共享交换和外部服务提供等过程进行统一管理。

4.4.4 数据资产价值管理用于实现数据资产的价值量化和运营，应具备以下能力：

- 1 支持数据资产评估，对数据资产进行质量评价和价值评估等，支持数据资产价值量化；
- 2 支持数据资产审计，监督数据资产管理和应用等过程的执行，识别风险，保障合规；
- 3 支持数据资产运营，对数据资产应用和服务进行计量计费 and 经营分析。

4.4.5 数据资产安全管理应具备以下能力：

- 1 支持数据资产脱敏，通过敏感数据规则设定，自动发现敏感数据并进行上报；
- 2 支持数据资产分级，可按部门和角色等多维度构建数据资产分级体系，实现数据资产的安全访问监管并支持按数据资产级别进行加密；
- 3 支持数据资产溯源，可通过数字水印、数据标识、区块链等技术，实现数据资产全生命周期可追溯。

4.5 数据质量管理技术要求

4.5.1 数据质量管理技术通过对不符合规则的数据进行探查、发现、修复、报告等，实现数据质量全流程闭环管理。宜包括数据质量探查、数据质量规则管理、数据质量评价、数据质量整改等技术能力。

4.5.2 数据质量探查应具备以下能力：

- 1 支持在无规则情况下对数据基础质量进行探查，对数据完整性、重复性、结构等进行检测；
- 2 支持对数据数量、空数据和重复数据进行统计分析；
- 3 支持基于元数据对数据存储位置、提供方式、总量、更新情况、数据结构等进行多维度探查。

4.5.3 数据质量规则管理应具备以下能力：

- 1 提供常用的数据质量模板，包括但不限于 SQL 规则、正则规则、值域规则、完整规则、重复规则、算法规则等；
- 2 支持数据质量的完整性、时效性、准确性、一致性、规范性、可访问性等多个维度的质量评价指标规则自定义配置；
- 3 支持基于数据标准的数据质量规则导入。

4.5.4 数据质量评价应具备以下能力：

- 1 支持对表级数据质量实施定期/临时稽核任务，发现表内不符合规则的数据，生成稽核结果文件；
- 2 提供数据质量事前、事中、事后监控，执行数据质量规则，发现数据质量趋势变化；
- 3 支持稽核结果采集和汇总操作；
- 4 支持自定义数据质量报告模板，提供多种组件、指标、维度的组合能力；
- 5 支持数据质量监控结果多方式告警。

4.5.5 数据质量整改应具备以下能力：

- 1 支持快速定位数据质量问题，包括但不限于问题数据的数量变化趋势分析、稽核类别分布分析、整体分析等手段；
- 2 支持对数据质量问题成因分析，支持对问题数据的查看、导出、血缘关系分析等；
- 3 支持对数据质量问题有效管理和跟踪，根据责任清单匹配数据责任人并派发问题，并对派发的数据质量问题进行跟踪管理、督办、评价等。

4.6 数据安全管理工作要求

4.6.1 数据安全管理工作通过数据访问控制、分类分级控制、脱敏加密、统计分析等技术手段，保障数据在存储、管理和使用等各个环节中的安全。应包括数据安全控制、数据安全保护、数据安全审计等技术能力。

【条文说明】数据安全管理工作要求符合《智慧民航数据治理 数据安全》（MH/T 5057）中的规定。

4.6.2 数据安全控制应具备以下能力：

- 1 支持数据访问授权，采取访问控制措施限制对数据源和数据存储系统的访问，对数据处理过程使用的工具，例如用于采集、脱敏、共享、管理等工具的访问、操作和运行进行权限控制；
- 2 支持身份鉴别，对登录的用户、进程/应用进行身份标识和鉴别，对系统中运行的数据

处理工具进行可信验证,防止工具假冒和篡改,对接入数据平台的采集/导出终端、数据导入/导出服务组件、分布式计算节点设备等进行身份鉴别;

3 支持分类分级控制,根据数据分类分级策略对数据进行标识和控制。

4.6.3 数据安全保护应针对数据的保密性、完整性和可用性提供全生命周期的防护,应具备以下能力:

1 支持数据脱敏,通过脱敏规则实现敏感数据的可靠保护;

2 支持数据加密,通过国家相关部门认可的加密算法对数据进行加密;

3 支持数据安全评估,对数据及其承载体(例如数据库等)的安全状况进行评估;

4 支持数据安全监测,利用数据流动监测、数据异常行为监测等实现对数据安全风险的全方位监测;

5 支持数据防泄露,包括但不限于以下方式:通过安全隔离技术实现批量数据泄露的网络拦截,通过数据脱敏实现外发敏感数据的匿名化,通过数据加密等技术实现核心数据加密存储;

6 支持数据安全标识,记录数据的安全属性,包括但不限于数据安全等级、业务类别、责任人等信息;

7 支持数据备份及数据损坏、丢失后的恢复;

8 支持数据溯源,记录原始数据在整个生命周期的演变信息和演变处理内容;

9 支持数据运维变更管控,防止数据误操作。

4.6.4 数据安全审计应具备以下能力:

1 支持数据安全事件信息统计,以数据安全分级策略为基础,对各类数据安全事件、告警等信息进行统计;

2 支持数据操作行为日志,对数据访问、数据操作等行为进行日志记录。

4.7 数据共享管理技术要求

4.7.1 数据共享管理技术通过提供多类型的数据共享方式,保障各节点业务系统之间数据交互。宜包括数据共享资源目录管理、数据共享服务、数据共享监控等技术能力。

4.7.2 数据共享资源目录管理应支持资源目录分类管理、资源编码、资源目录维护、资源目录检索、批量服务、实时服务等。

4.7.3 数据共享服务应具备以下能力:

1 支持数据抽取、库表同步、文件同步、消息订阅推送等方式的批量数据服务;

2 支持单表、结果集方式的接口数据服务;

3 支持通过可视化界面进行服务编排、配置、规则定义及服务接口发布;

- 4 支持数据使用方通过资源申请或订阅的方式获取数据资源；
 - 5 支持通过服务授权机制对数据使用方提出的数据使用申请进行服务授权；
 - 6 支持关键共享数据对账机制，确保共享数据能够准确传输。
- 4.7.4** 数据共享监控应支持对数据、文件传输和共享服务运行状态进行监控，当数据共享失败时自动触发故障告警。

5 民航数据操作类技术要求

5.1 数据采集技术要求

5.1.1 数据采集技术用于从不同数据源获取数据，并对数据进行标准化处理以满足应用需求。宜包括数据接入、数据获取、条件过滤、作业调度、数据清洗等技术能力。

5.1.2 数据接入宜支持对主流数据库、文件数据源、互联网数据、物联网数据、接口数据源、文件传输代理、前置交换等多种形态数据源的接入，并对接口、任务以及接入的数据项数量和完整性等进行监控。

5.1.3 数据获取应支持全量/历史数据获取和实时/增量数据获取，具体要求如下：

1 全量/历史数据获取应支持数据传输服务、高并发的离线数据上传下载服务，支持全量数据或历史数据的批量导入及导出；

2 实时/增量获取应支持实时同步、定时获取、数据订阅、日志获取等服务。

5.1.4 条件过滤应支持按照指定条件进行指定过滤采集（例如根据字段过滤等）。

5.1.5 作业调度应支持条件触发、并发调度、周期循环调度等模式，支持对作业启动、停止、暂停、恢复等操作。

5.1.6 数据清洗宜包含数据质量分析、定义清洗规则、清洗数据、清洗结果验证等能力，具体要求如下：

1 数据质量分析应支持对数据源进行分析，及时发现数据源存在的质量问题；

2 定义清洗规则应支持对空值、非法值、不一致数据、相似重复记录等数据的检测和处理；

3 清洗数据应根据规则，补足残缺和空值、纠正不一致、进行数据拆分、合并或去重、数据脱敏、数据除噪等；

4 清洗结果验证应支持对清洗方法的正确性和效率进行验证与评估，对不满足要求的清洗方法进行调整和改进。数据清洗过程宜多次迭代并进行分析、设计和验证。

5.2 数据存储技术要求

5.2.1 数据存储技术用于满足大规模、多元化的数据管理需求，支持数据治理全流程的数据存储高性能、高可用。宜包括结构化数据存储、列式数据存储、图数据存储、文件存储、数据备份等技术能力。

【条文说明】 数据存储包括关系型数据库存储和 NoSQL 数据库存储，支持存储视频、音频、电子文档、日志等半结构化、非结构化数据，支持流数据存储。

5.2.2 结构化数据存储符合下列要求：

- 1 应支持结构化数据的分布式存储机制，实现数据存储的可扩展性；
- 2 应支持 API 接口实现数据的各类查询操作；
- 3 应支持多表关联；
- 4 应支持数据分布式存储的一致性；
- 5 宜支持行列混合存储和行列转换。

5.2.3 列式数据存储符合下列要求：

- 1 应支持数据以键值形式进行存储；
- 2 应支持基于表、列族和列的用户权限管理；
- 3 应支持列加密；
- 4 宜支持多级索引；
- 5 宜支持将多个具有类似功能或存在关联的业务表合并存储。

5.2.4 图数据存储符合下列要求：

- 1 应支持由节点及边组成的数据模型；
- 2 应支持图查询、图遍历及图分析；
- 3 应支持最短路径、最优路径遍历搜索；
- 4 宜支持顶点、属性的继承操作；
- 5 宜支持长任务异步会话机制。

5.2.5 文件存储符合下列要求：

- 1 应支持文件的上传、下载、读写、复制、移动、删除、访问控制；
- 2 应支持文件的容错机制和系统的高可用机制，包括数据块的备份、系统快速恢复等；
- 3 应支持文件数据的校验和同步，保证数据的完整性与一致性；
- 4 应支持分布式的弹性扩展，支持动态添加以及删除节点；
- 5 应支持文件的搜索、批量操作、回收站、快照等能力。

5.2.6 数据备份应支持针对不同数据类型和安全等级,采用不同的备份机制保障数据完整性、可用性与安全性。

5.3 数据计算技术要求

5.3.1 数据计算技术用于满足数据治理过程中不同业务、不同量级的数据加工和计算需求,为数据融合和分析提供计算引擎。宜包括批处理、流处理、图处理、MPP 交互分析等技术能力。

5.3.2 批处理符合下列要求:

- 1 应支持多种数据类型的离线分析,包括结构化、非结构化数据;
- 2 应支持离线计算任务进度与状态的实时上报;
- 3 应支持多节点离线任务联动执行;
- 4 应支持分散-聚集的处理方式;
- 5 应支持批处理计算框架在分布式资源上运行。

5.3.3 流处理符合下列要求:

- 1 应支持从数据源中获得实时消息数据,完成高吞吐、低延迟的实时计算,并将结果输出到消息队列或者进行持久化;
- 2 应支持用户级别的访问控制,支持对消息处理任务进行创建、浏览、中止、激活等操作,并记录操作日志;
- 3 宜支持高容错,当处理进程出现异常时,能够重新部署该处理单元。

【条文说明】流处理适用于时效性要求较高的场景,例如实时推荐、业务监控和预警等。

5.3.4 图处理符合下列要求:

- 1 应支持图数据查询类的 API,支持同步计算模型或异步计算模型编写迭代算法;
- 2 应支持基于属性图模型的图数据表达,包含结点/边上的标签和属性类型定义;
- 3 应支持常用图指标计算,包含描述图的拓扑结构特征等。

5.3.5 MPP 交互分析应支持采用大规模并行计算架构,利用集群多节点的计算能力和内存提升总体性能,支持复杂的多表连接计算。

5.4 数据融合技术要求

5.4.1 数据融合技术用于打通不同业务、不同数据之间的壁垒,融合多个数据源产生新的信息,支撑一体化数据分析。宜包括数据汇聚、数据主题联接等技术能力。

5.4.2 数据汇聚指数据从数据源通过响应请求或主动推送的方式进入数据资源池。宜包括物理汇聚和虚拟汇聚能力，具体要求如下：

1 物理汇聚：应支持将原始数据复制到数据资源池中，包括但不限于批量集成、数据复制同步、消息集成、流集成等；

2 虚拟汇聚：应支持建立对应虚拟表的集成方式实现汇聚，原始数据不在数据资源池中进行物理存储。

5.4.3 数据主题联接指在数据资源池的基础上建立数据联接层，基于不同分析场景，通过多种方式将跨域数据建立联接。宜包含多维模型、图模型、标签、指标、算法模型等能力，具体要求如下：

1 多维模型：应支持依据明确的业务关系，建立基于维度、事实表以及相互间连接关系的模型，实现多维度、多层次的数据分析服务；

【条文说明】例如，对航空公司订票数据从时间、区域、产品、旅客等维度进行多视角、不同粒度的查询和分析。

2 图模型：应支持建立数据对象以及数据实例之间的关系，进行数据间的关联影响分析；

【条文说明】例如，查看机场的通航点数据具体关联到哪个航空公司以及时刻、机型等详细信息时，可以通过图模型快速分析关联影响，支撑业务决策。

3 标签：应支持根据业务场景的需求对目标对象运用抽象、归纳、推理等算法得到的高度精练的特征标识，进行高效的数据筛选和分析；

4 指标：应支持依据明确的业务规则，通过数据计算得到衡量目标总体特征的统计数值，度量业务结果、效率和质量，客观表征业务状况；

5 算法模型：应支持根据业务需求，运用数学建模对现实世界进行抽象、模拟和优化。

5.5 数据分析技术要求

5.5.1 数据分析技术用于数据业务价值的挖掘和最终呈现，为数据决策提供支撑。宜支持多种数据分析模式，宜支持可视化分析、统计分析、机器学习等技术能力。

【条文说明】例如，通过对航班 QAR（快速存储记录器）数据进行分析，实现航空公司运行品质安全水平综合评估预测，对实际运行风险防控起到积极作用。

5.5.2 数据分析模式应支持离线数据分析模式、流数据分析模式、交互式联机分析模式等，具体要求如下：

1 离线数据分析模式符合下列要求：

- 1) 应支持对离线数据的分布式分析；
- 2) 应支持具有标准接口的第三方应用；

- 3) 应支持分布式计算或并行计算等计算框架;
- 4) 应支持对工作任务的切分和分布式调度;
- 5) 宜支持对关系型数据库和大数据存储系统中的数据进行交叉查询、聚合、关联操作。

2 流数据分析模式符合下列要求:

- 1) 应支持按时间切片后进行批量处理;
- 2) 应支持基于事件触发或者采样的流式处理;
- 3) 应支持实时流上的数据统计;
- 4) 应支持流式数据的排序;
- 5) 应支持多个数据流的关联处理。

3 交互式联机分析模式符合下列要求:

- 1) 应支持通过结构化查询语言对数据进行分布式的联机分析, 例如 OLAP 等;
- 2) 应支持通过结构化查询语言对数据进行即席查询;
- 3) 应支持利用可视化中间件对数据分析结果进行展示;
- 4) 应支持基于在线联机分析的交互式数据分析;
- 5) 宜支持对非结构化数据的分析。

5.5.3 可视化分析应支持高维数据静态和动态可视化展示。

5.5.4 统计分析应支持数值统计分析、集中趋势统计分析、离散程度统计分析等典型统计分析应用, 以及制定自定义模板以保存常用的统计分析方案。

5.5.5 机器学习能力符合下列要求:

- 1 应支持将输入数据划分为训练集、验证集和测试集;
- 2 应支持对视频等非结构化数据的分析;
- 3 应支持机器学习模型的导入、校验和导出;
- 4 宜提供机器学习的流程组件, 包括特征提取、特征转换、特征选择、模型选择、交叉验证、模型调优等;
- 5 宜支持多种机器学习算法, 包括但不限于回归与分类算法、聚类算法、降维算法、频繁模式挖掘算法、神经网络算法等。

标准用词说明

1 为了便于在执行本规范条文时区别对待，对要求严格程度不同的用词，说明如下：

1) 表示很严格，非这样做不可的用词：

正面词采用“必须”；反面词采用“严禁”。

2) 表示严格，在正常情况下均应这样做的用词：

正面词采用“应”；反面词采用“不应”或“不得”。

3) 表示允许稍有选择，在条件许可时首先这样做的用词：

正面词采用“宜”；反面词采用“不宜”。

4) 表示有选择，在一定条件下可以这样做的，采用“可”。

2 本规范中指定应按其他有关标准、规范执行时，写法为“应符合……的规定”或“应按……的规定执行”。非必须按所指定的标准、规范和其他规定执行时，写法为“可参照……”。

引用标准名录

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，仅所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（包含所有修改单）适用于本文件。

- [1] 《信息技术服务 治理 第5部分：数据治理规范》(GB/T 34960.5)
- [2] 《数据管理能力成熟度评估模型》(GB/T 36073)
- [3] 《信息技术 大数据分析系统功能要求》(GB/T 37721)
- [4] 《信息技术 大数据存储与处理系统功能要求》(GB/T 37722)
- [5] 《信息安全技术 数据安全能力成熟度模型》(GB/T 37988)
- [6] 《智慧民航数据治理规范 框架与管理机制》(MH/T 5054)
- [7] 《智慧民航数据治理规范 数据架构》(MH/T 5055)
- [8] 《智慧民航数据治理规范 数据质量》(MH/T 5056)
- [9] 《智慧民航数据治理规范 数据安全》(MH/T 5057)
- [10] 《智慧民航数据治理规范 数据服务》(MH/T 5058)