

直升机大数据治理系统架构设计研究

鲁兴¹, 孟浩¹, 吕少杰², 张京亮²

(1. 中国直升机设计研究所, 江西 景德镇 333000; 2. 中国人民解放军 66037 部队, 北京 101121)

摘要: 高质量数据是模型算法迭代更新的先决条件, 是实现直升机高质量发展的基础。直升机研制和使用过程中产生了大量技术与运用数据, 但存在“有数据、无治理”、数据缺乏统一管理、价值挖掘不足等问题。结合直升机领域机械化、信息化、智能化转型发展需求, 对直升机大数据治理体系进行了探讨, 提出了基于云边端协同的直升机大数据治理架构, 并详细分析了直升机湖仓一体数据云总体建设思路, 为构建直升机领域数据治理体系提供技术支持, 奠定了直升机“+智能”的数据高质量基础。

关键词: 直升机; 大数据治理系统; 云边端协同; 湖仓一体数据云

中图分类号: TP399; V24

文献标志码: A

DOI: 10.19358/j.issn.2097-1788.2026.02.008

中文引用格式: 鲁兴, 孟浩, 吕少杰, 等. 直升机大数据治理系统架构设计研究 [J]. 网络安全与数据治理, 2026, 45(2): 60-64.

英文引用格式: Lu Xing, Meng Hao, Lv Shaojie, et al. Research on architecture construction of helicopter big data governance system [J]. Cyber Security and Data Governance, 2026, 45(2): 60-64.

Research on architecture construction of helicopter big data governance system

Lu Xing¹, Meng Hao¹, Lv Shaojie², Zhang Jingliang²

(1. China Helicopter Research and Development Institute, Jingdezhen 333000, China;

2. Unit 66037 of the PLA, Beijing 101121, China)

Abstract: High-quality data is a prerequisite for the iterative updating of model algorithms and serves as the foundation for achieving optimal development in the field of helicopters. Throughout the development process and usage life of helicopters, a significant amount of technical and application data is generated. However, challenges such as unregulated data generation, lack of unified data management, and inadequate value extraction persist are still huge application barriers. To address these issues and meet the requirements posed by mechanized construction, information technology integration, and intelligent transformation in the helicopter industry, a helicopter big data governance system was proposed. A cloud-side-end collaboration-driven helicopter big data governance architecture was illustrated. The overall construction idea of helicopter lake-stash integrated data cloud was analyzed in detail. The proposed governance system offers essential technical support for establishing effective data governance practices within the helicopter domain while laying the foundation for enhancing intelligence through high-quality helicopter-related datasets.

Key words: helicopter; big data governance system; cloud-side-end collaboration-driven; lake-stash integrated data cloud

0 引言

随着科技的发展, 直升机行业先后经历了机械化、信息化, 现阶段正在向智能化演进。机械化的主导要素是机动力和操作性, 核心在于直升机自动化能力提升; 信息化的主导要素是信息力, 制信息权成为本领域的核心要素; 未来直升机智能化的主导要素是智能优势, 包括自主智能、智能决策等, “制智权”成为行业领域的核心。

数据是智能化发展的“燃料”, 是夺取“制智权”

的核心基础支撑。我国关于科技创新的规划决策中指出, 我国大数据采集和应用刚刚起步, 要加强研究, 加大投入, 力争走在世界前列^[1]。大力推进直升机领域大数据相关问题研究, 是贯彻落实国家科技创新战略部署的具体举措。2012年6月, 我国正式设立了专职责任机构并成立了数据中心和保障队, 2023年2月, 中共中央、国务院印发了《数字中国建设整体布局规划》, 其中指出建设数字中国是数字时代推进中国式现代化的重要引擎^[2-3]。

直升机是交通运输领域的重要组成部分,在直升机研制及使用过程当中,会源源不断地产生极为庞大的技术与运用数据。通过构建直升机飞行数据综合管理系统与发动机传动健康监测与故障诊断系统^[4-6],初步建立了涵盖直升机日常飞行、训练、维护和保障等方面的数据采集与治理体系框架。采集的数据涵盖了直升机的性能参数、运行状态、使用效果等多个方面,具有极高的价值。但直升机数据治理领域仍存在“有数据、无治理”的突出问题,这些海量的数据因缺乏统一的管理模式而无法充分发挥其价值^[7],具体表现在以下两个方面。一方面,大量的数据分散在各个不同的系统和部门中,如直升机飞行数据综合管理系统与发动机传动健康监测与故障诊断系统等多个不同的数据采集系统中,没有形成一个集中、统一的数据存储和管理平台。这使得数据的查询、调用和分析变得极为不易,降低了数据的利用效率。另一方面,由于缺乏有效的数据治理机制,各个数据采集与处理系统中数据格式、类型、采样率、采样时间等均不一致,导致数据的准确性、完整性和一致性难以得到保证。数据中的错误、缺失和不一致性,数据标准和规范不统一,导致不同系统之间的数据难以互通共享,阻碍了数据的流通和整合^[8-9]。

本文的创新点在于,针对现有直升机数据治理领域缺乏统一的数据存储、管理和数据治理机制的问题,对照航空产品建设目标中关于实现行业数字化“全覆盖”、数据治理流程“全打通”的要求^[10],构建基于“云-边-端”的直升机领域数据治理体系,打通典型任务流程和要素的数据架构标准和治理流程,开发数据集成、数据生成、数据分析处理等数据治理工具链,建成覆盖典型场景和要素的数据资源池,初步形成陆航领域数据架构标准、基础数据平台和数据治理工具链^[11-14]。

1 直升机数据治理云边端协同系统总体架构

数据治理是算法迭代演进的基础支撑,本文旨在研究建立直升机数据治理系统架构,包括数据采集层、数据处理层和应用服务层,总体架构如图1所示。通过云边端下的信息通信协同、数据存储协同和计算服务协同实现对直升机领域数据的治理,满足算法模型高效训练、快速迭代的数据需求^[15]。

终端采集设备是直升机大数据的来源,负责监控并收集各种类型的数据,具体包括设备类型、编号、生产信息、飞行状态、飞行参数、振动/载荷数据、雷达数据、光电数据等。在云边端协同架构中,设备终端依托统一采集通信模块实现数据采集汇聚的功能,

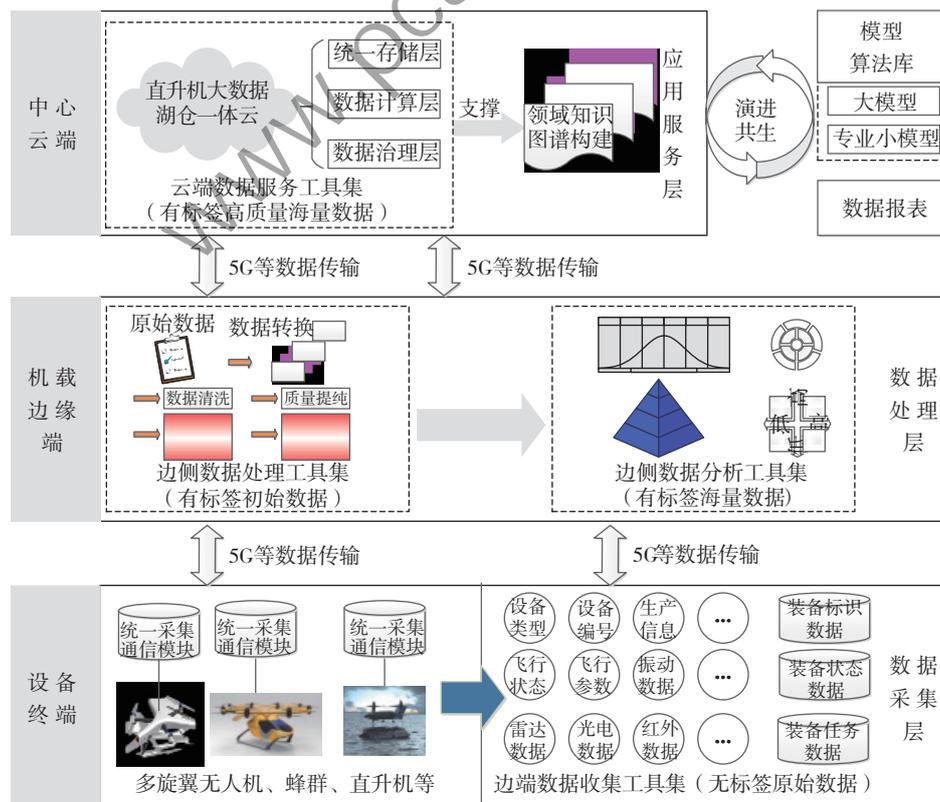


图1 云边端协同的直升机大数据治理架构

基于 5G 网络将实时数据传输到机载边缘计算平台,进行更高质量的数据处理与分析,为直升机云边端协同架构提供了坚实的基础。

机载边缘端负责对收集到的直升机数据进行预处理,首先,构建统一的数据治理体系标准与规范,根据标准规范将不同的数据转换成统一格式;其次,通过缺失值处理、异常过滤、噪声清除、数据去重、错误去重等操作,完成数据的清洗与质量提纯;再次,依托数据分析工具集对数据进行智能分析,通过聚类、相关性检测等方法挑选出最有用的数据,通过局域网/5G 等传输方式传输到数据中心云端,为直升机载应用服务提供数据支撑。

中心云端直升机大数据湖仓一体云为直升机应用服务提供了大规模的计算和存储资源,持续引接汇聚上传的数据,通过海量高质量数据构建直升机领域知识图谱,形成相关知识库,满足专业模型算法库对多模态、高质量数据的需求,并将结果返回给边缘计算设备与端侧执行机构。

依托构建的云边端协同数据治理体系,实现对数据的全方位治理,得到全面、高质量的直升机领域数据。依托大数据联机分析、实时分析功能,可充分发挥数据的作用,加快单体算法、群体算法、专用小模型、通用大模型的迭代演进速度。

2 直升机湖仓一体云建设思路

该部分对直升机大数据湖仓一体架构进行详细的说明。大数据湖仓一体是将数据仓库和数据湖的优势融合在一起的一种数据管理架构。传统的数据仓库主要用于存储结构化数据,支持快速查询和分析;而数据湖则用于存储大量原始的、非结构化的或半结构化的数据,提供更大的灵活性和扩展性。大数据湖仓一体通过整合这两种技术,实现了数据的高效管理和分析。大模型需要各种形态(如数据表、文档、图片

等)的数据,而湖仓一体架构很好地满足了大模型对多模态、高质量数据的需求。湖仓一体基本架构图如图 2 所示,主要由以下三个关键组件构成。

统一存储层:统一存储层是湖仓一体架构的核心组成部分,负责存储各种类型的数据,包括原始数据、加工清洗数据、模型化数据等。

数据计算层:数据计算层负责数据处理,可集成多种计算引擎和工具,支持从简单的数据查询到复杂的数据分析。此外,用户可以对存储在数据湖中的数据进行实时或批量的处理和分析,以满足大模型场景下各种数据需求。

数据治理层:数据治理主要包括数据接入、元数据管理、数据质量管理、数据安全等功能。通过数据治理可以在确保数据安全的情况下,高效便捷地使用高质量数据,从而让数据发挥更大的价值。

2.1 统一存储层

湖仓一体架构将数据仓库和数据湖集成在一起,为用户提供统一的数据存储层。技术实现上主要由云存储、开放的文件格式(如 Parquet、ORC 等)和开放的表格式(如 Iceberg、Hudi、Delta Lake 等)组成。

(1) 云存储:提供高可用性、持久性和可扩展性的存储服务,通过使用开源的 Minio 对象存储或者本地 HDFS 存储,云存储实现了存储与计算的分离。

(2) 开放的文件格式:采用如 CSV、JSON、XML 等常用文件格式在云存储中存储不同类型的数据。对于 OLAP 分析,采用 Parquet、ORC 等列式存储格式,提供更高的查询性能和压缩比。

(3) 开放的表格式:开放表格式(常见的如 Iceberg、Hudi 和 Delta Lake 等)提供对数据读写相关的事务支持、时间回溯、SQL 查询等功能,使数据在湖仓一体架构中灵活应用,便于管理。

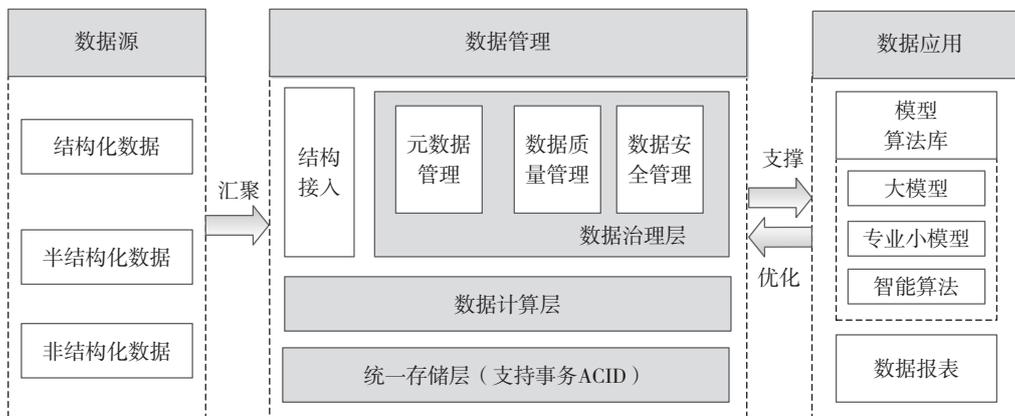


图 2 湖仓一体基本架构图

2.2 数据计算层

数据计算层主要处理分析存储在湖上的数据。从技术架构上可以支持多种计算引擎（如 Spark、Flink、StarRocks 等），这些引擎可以直接访问或查询存储层中的数据，实现高效的数据处理和分析。虽然数据分析场景和需求各式各样，但从数据需求时效性看主要分为离线计算和实时计算。

(1) 离线批处理计算：主要用于处理大量静态数据。主要完成直升机设计、制造、飞行等全生命周期内原始数据的过滤、清洗、提纯、计算等，从产生的海量、原始数据中初步获取数据中的价值信息，筛选出高价值数据，完成分类筛选预处理。这种处理模式适用于对实时性要求不高的场景，一般处理数据量较大。

(2) 实时流处理计算：主要用于处理实时数据流。对实际飞行过程中产生的飞行、振动、载荷等数据进行快速、实时的处理和分析，及时发现飞行过程中的异常状态，快速完成飞行过程健康监控与故障诊断，及时为飞行安全提供在线的反馈和预警。这种处理模式适用于需要快速响应的场景，一般都会涉及飞行品质与飞行安全。

2.3 数据治理层

数据治理层主要利用数理统计、深度挖掘等方法，从数据库中寻找有用的知识，得到高质量的数据集。该层重点完成数据元数据管理、数据质量管理、数据安全等内容，经过数据治理、加工过程，满足不同任务场景下的大小模型等算法库的训练需求。

(1) 元数据管理：本质是有效利用数据资产，让数据发挥出尽可能大的价值。统一的元数据管理可以有效打破数据孤岛现象，实现数据的统一管理和共享。元数据管理一般包括元模型管理、元数据审核、元数据维护、元数据版本管理、元数据变更管理等内容。

(2) 数据质量管理：目标是确保数据准确、一致和完整，这对大模型的效果有非常大的影响，如高质量的数据可以降低大模型的幻觉。提升数据质量主要包括数据标注、数据清洗、数据校验以及数据去重等手段。

①数据标注技术：组织专业人员进行数据初步标注，建立数据标注标准、规范，提升标注准确性。利用大模型的强大学习能力和预测能力，对数据进行初步标注，然后再由专业人员进行审核和修正，实现机器性能快速迭代、数据自动分级分发与渐进式辅助标注。通过人机耦合持续演进，不断提升数据标注效率，

降低标注成本。

②数据清洗技术：对数据进行处理和修正，以消除数据中的错误、重复、不完整或不准确的信息，常见的处理技术包括缺失值处理、纠正错误记录、格式转换、数据合并、数据规范化和数据归一化等^[16-17]。图3所示为数据清洗逻辑框图。

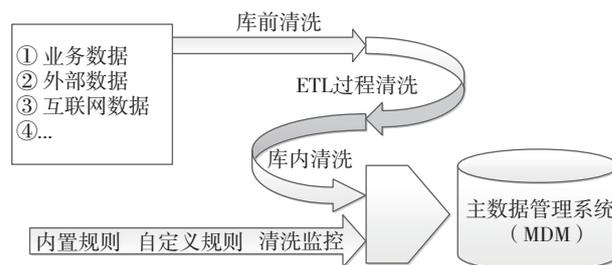


图3 数据清洗逻辑图

③数据效验技术：对数据进行验证和检查，以确保数据的准确性、完整性和一致性。主要包括格式验证、范围验证、空值验证、数据类型验证、唯一性验证、关联性验证、数据完整性验证等。

④数据去重技术：从数据仓库中去除重复的数据，减少数据的冗余和提高数据的质量，可以综合采用的方法包括基于比较的去重、基于特征的去重、基于聚类的去重、基于关联规则挖掘的去重等。

(3) 数据安全治理：基于湖仓一体的统一存储，可以让数据在更大范围内得到共享和使用，但会对数据的安全性带来挑战。为此，可通过采取包括数据访问前进行用户身份认证、数据访问时进行鉴权和访问控制、访问后进行审计等措施确保数据安全。

3 结论

随着直升机机械化、信息化、智能化融合发展需求，直升机设计、制造、使用、维护等全生命周期管理过程中的数据量和数据复杂性急剧增加，对直升机领域数据持续引接汇聚、数据存储和数据应用服务提出了更高的要求。本文以直升机领域大数据治理为核心问题，构建了基于云边端协同的直升机大数据治理架构，从数据治理、数据计算、数据存储的维度分析了直升机湖仓一体云构建关键技术，形成直升机领域数据资源池，为实现直升机智能化应用发展提供重要支撑。

参考文献

- [1] 中共中央文献研究室. 习近平关于科技创新论述摘编 [M]. 北京: 中央文献出版社, 2016.
- [2] 郑少秋, 韩立斌, 王静, 等. 军事大数据概念内涵、发展挑

战与技术实践 [J]. 指挥与控制学报, 2023, 9 (5): 508 - 516.

[3] 蒋猛, 禹明刚, 王智学. 对我军大数据元数据构建问题的思考 [J]. 指挥与控制学报, 2018, 4 (3): 220 - 225.

[4] 于海龙. 直升机飞行数据综合管理系统分析 [J]. 军民两用技术与产品, 2016 (4): 20, 30.

[5] 黄润龙. 新一代直升机综合数据处理平台设计及应用 [J]. 电讯技术, 2019, 59 (10): 1163 - 1168.

[6] 李昊晗, 杨春宝, 时丽, 等. 直升机发动机故障诊断及控制策略 [C]//2023 年中国航空工业技术装备工程协会年会, 2023: 141 - 143.

[7] 贲婷婷, 杨柳静, 黄太奇. 面向数字化转型的军事数据治理通用实施路径 [J]. 指挥信息系统与技术, 2023, 14 (2): 23 - 27, 68.

[8] 董鸿鹏, 彭紫薇, 张俊. 航空兵任务规划系统数据组织管理 [J]. 指挥信息系统与技术, 2023, 14 (2): 87 - 90.

[9] 董鸿鹏, 金雷, 彭紫薇. 航空兵任务规划系统数据库设计 [J]. 兵工自动化, 2023, 42 (4): 17 - 20, 35.

[10] 甘牛, 李熙, 张莹莹, 等. 航天装备制造企业数据治理体系建设与实践 [J]. 中国高新科技, 2023 (13): 153 - 157.

[11] 钟正仪, 包卫东, 王吉, 等. 一种面向云边端系统的分层异构联邦学习方法 [J]. 计算机研究与发展, 2022, 59 (11): 2408 - 2422.

[12] 李敏, 马帅印, 殷磊, 等. 云边端协同驱动的陶瓷制造过程能效调度方法 [J]. 兵器装备工程学报, 2024, 45 (8): 174 - 183.

[13] 戴宁, 徐开心, 胡旭东, 等. 云边端协同下的织机设备信息采集与监测系统 [J]. 纺织学报, 2024, 45 (8): 215 - 224.

[14] 李高云, 旷生玉, 江果, 等. 电磁大数据云边端协同应用架构研究 [J]. 航天电子对抗, 2022, 38 (4): 4 - 8, 19.

[15] 龚佳, 谢全, 王圣东. 海军作战大数据处理系统建设及应用 [J]. 指挥信息系统与技术, 2015, 6 (2): 22 - 26.

[16] MA S Y, HUANG Y M, LIU Y, et al. Big data-driven correlation analysis based on clustering for energy-intensive manufacturing industries [J]. Applied Energy, 2023, 349: 121608.

[17] QU X F, WANG H Q. Emergency task offloading strategy based on cloud-edge-end collaboration for smart factories [J]. Computer Networks, 2023, 234: 109915.

(收稿日期: 2025 - 06 - 29)

作者简介:

鲁兴 (1989 -), 通信作者, 男, 博士, 工程师, 主要研究方向: 自动化、智能算法。E-mail: luxing0131@hotmail.com。

孟浩 (1989 -), 男, 硕士, 高级工程师, 主要研究方向: 计算机、大数据。

吕少杰 (1983 -), 男, 博士, 高级工程师, 主要研究方向: 直升机技术、计算机。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com