

基于数据中台的智慧校园数据体系架构设计

李楠楠¹, 王振锟¹, 陈青华¹, 樊辉锦²

1. 海军航空大学岸防兵学院, 山东 烟台 264000; 2. 海军指挥学院研究生大队, 南京 210018

【摘要】为支撑智慧校园建设,结合行业最新发展趋势,参考主流高校的成功经验,开展基于数据中台的智慧校园数据体系设计研究。文章提出基于数据中台技术架构的智慧校园实现方式,描述开展数据治理的主要内容、实施步骤,实现智慧校园业务数据统一采集、数据资产统一管理、数据服务统一发布,为上层应用系统提供数据查询、数据分析等服务;同时,分析升级架构前后的数据综合运用情况。结果表明,在智慧校园中应用数据中台架构能够形成“顶用、好用”的数据生态圈,有利于助推学校转型升级,提高人才培养质量。

【关键词】智慧校园;数据中台;数据治理;

【中图分类号】G434 **【文献标志码】**A **【文章编号】**1004-5287(2023)06-0694-06

【DOI】10.13566/j.cnki.cmet.cn61-1317/g4.202306013

Smart campus data system design based on data middle platform

LI Nannan¹, WANG Zhenkun¹, CHEN Qinghua¹, FAN Huijin²

1. College of Coastal Defense, Naval Aviation University, Yantai 264000, China;

2. Postgraduate Brigade, Naval Command College, Nanjing 210018, China

【Abstract】In order to support the construction of smart campus, the research on the design of data system of smart campus based on data middle platform is carried out by combining the latest development trend of the industry and referring to the successful experience of mainstream universities. The paper proposes the realization method of smart campus based on data middle platform technology, describes the main contents and implementation steps of data governance, realises unified collection of business data, unified management of data assets and unified release of data services for the smart campus, and provides data query and data analysis services for the upper-level application systems. Meanwhile, comprehensive application of data before and after construction upgrading is analyzed. The results show that the application of the data middle platform architecture in the smart campus can form a data ecosystem that is “useful and useful”, which is conducive to the transformation and upgrading of the university and improving the quality of talent training.

【Key words】smart campus; data middle platform; data governance

为了支撑高校智慧校园建设,基于数据中台技术能够助力高校从传统 IT 架构向微服务架构转型。通过盘点数据资产、制定数据标准和管理规范,加快数据资源的汇聚,建设完善统一的数据资产目录,能够逐步形成覆盖高校各业务的大数据体系;同时,实现数据资源的有效整合、挖掘分析、统一管理和资源共享,赋能各传统业务系统,为高校精细化管理和辅助决策提供数据支撑^[1]。结合研究文献^[2-4]

对数据中台技术的研究与运用,文章认为数据中台技术已经不再是狭义上阿里公司出品的一个大数据服务平台,而是以大数据服务为基础,包含数据采集、管理、分析和治理等功能,同时能够支撑业务应用标准化和定制化的低冗余、高容错、可复用的数据服务管理平台。文章主要从体系架构设计和实例统计分析两个方面,对基于数据中台的智慧校园数据体系设计展开研究。

1 体系架构

智慧校园体系架构的数据底座采用数据中台进行打造,既可以提供数据汇聚、数据开发、数据资产管理、数据服务、数据可视化等功能,也能够有效解决各种校园应用之间的数据共享与交换,以支撑学

校各部门的业务协同联动。同时,数据中台可以屏蔽底层计算存储平台的技术复杂性,方便运维人员快速、简便地实现数据采集、数据分析、数据管理等操作,降低学校对数据技术人才的依赖性,让数据的使用成本更低^[5-6]。



图 1 基于数据中台的智慧校园体系架构

1.1 数据中台

数据中台在智慧校园数据体系中扮演着核心重要角色,可为智慧校园提供数据汇聚、数据开发、数据资产管理、数据治理和数据服务等业务。同时,为智慧校园的数据挖掘和数据分析提供支持,向下通

过数据汇集将智慧校园的业务数据集中沉淀到数据中台,实现数据在多个业务系统之间的数据交互和采集;向上为顶层智慧应用提供数据支撑,推动智慧校园数据业务高效运转的同时充分发挥数据价值。数据中台的主要框架如图 2 所示。

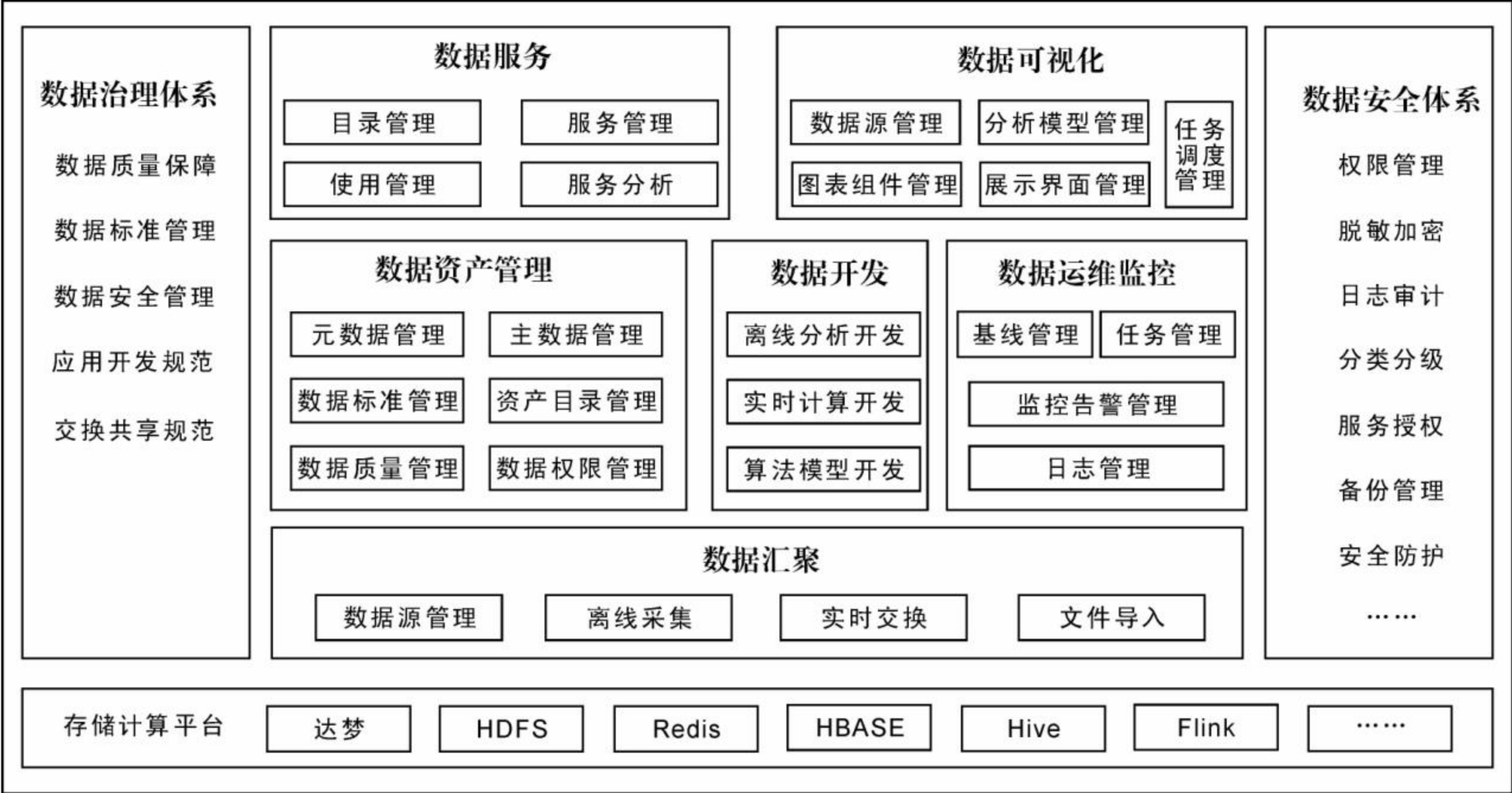


图 2 数据中台架构

1.1.1 数据汇聚

数据汇聚是数据中台的接入口,所有数据来自于学校各业务系统、日志、文件、网络等。这些数据分散在不同的网系和应用存储中,若仅被自身业务环境利用,所产生的数据价值很低。通过数据汇聚将分散的资源进行集成整合,才能实现全面的共享。作为数据中台的核心功能,数据汇聚将各种网系、各类异构的数据便捷地采集到同一平台进行集中存

储,为后续的加工建模做准备,主要用于校园跨部门、跨层级、跨网络之间的安全采集、数据传输任务,并提供数据管理的流程审计、过程跟踪、内容过滤等功能。数据汇聚主要包括数据源管理、离线采集、实时交换和文件导入等,能够支持多源数据库及文件的采集和预处理,满足多种校园应用场景的需求。在实现方面,数据汇聚架构如图 3 所示。通过各类业务系统配备的底层数据采集工具以及人员填

报信息的相关操作,使用固定的数据传输协议,例如 MQTT 协议、HTTP 协议等,将所采集到的基础业务信息数据发送至数据队列;数据队列对数据进行分类处理,区分实时数据与非实时数据,非实时数据存入存储计算平台,实时数据交由实时分布式数据库进行处理;也可以通过离线采集和文件导入的方式完成数据采集与汇聚工作。同时,用户需要调用或订阅数据时,可以依托 Spark Streaming 等流处理集群进行大数据处理以提供相应服务。

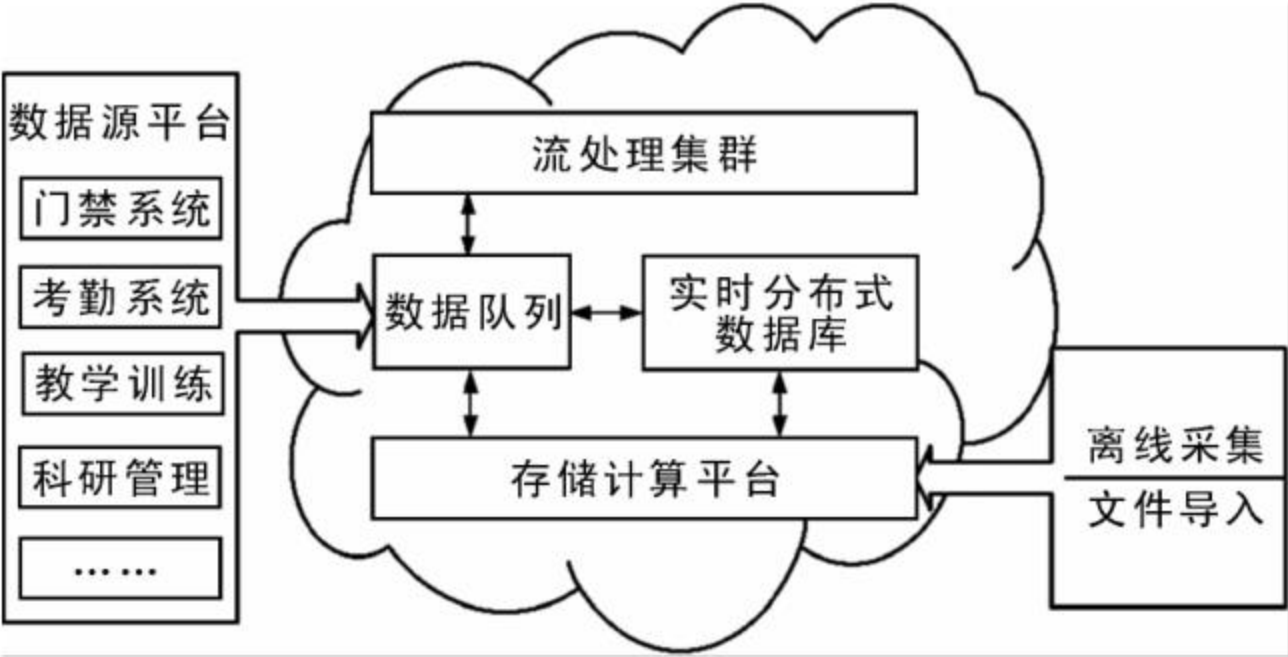


图 3 数据汇聚架构

1.1.2 数据模型

数据中台的数据模型层次可以分为三层,分别是数据应用层(application data service,ADS)、公共维度层(common data model,CDM)和数据操作层(operation data store,ODS)。主要模型架构如图 4 所示。

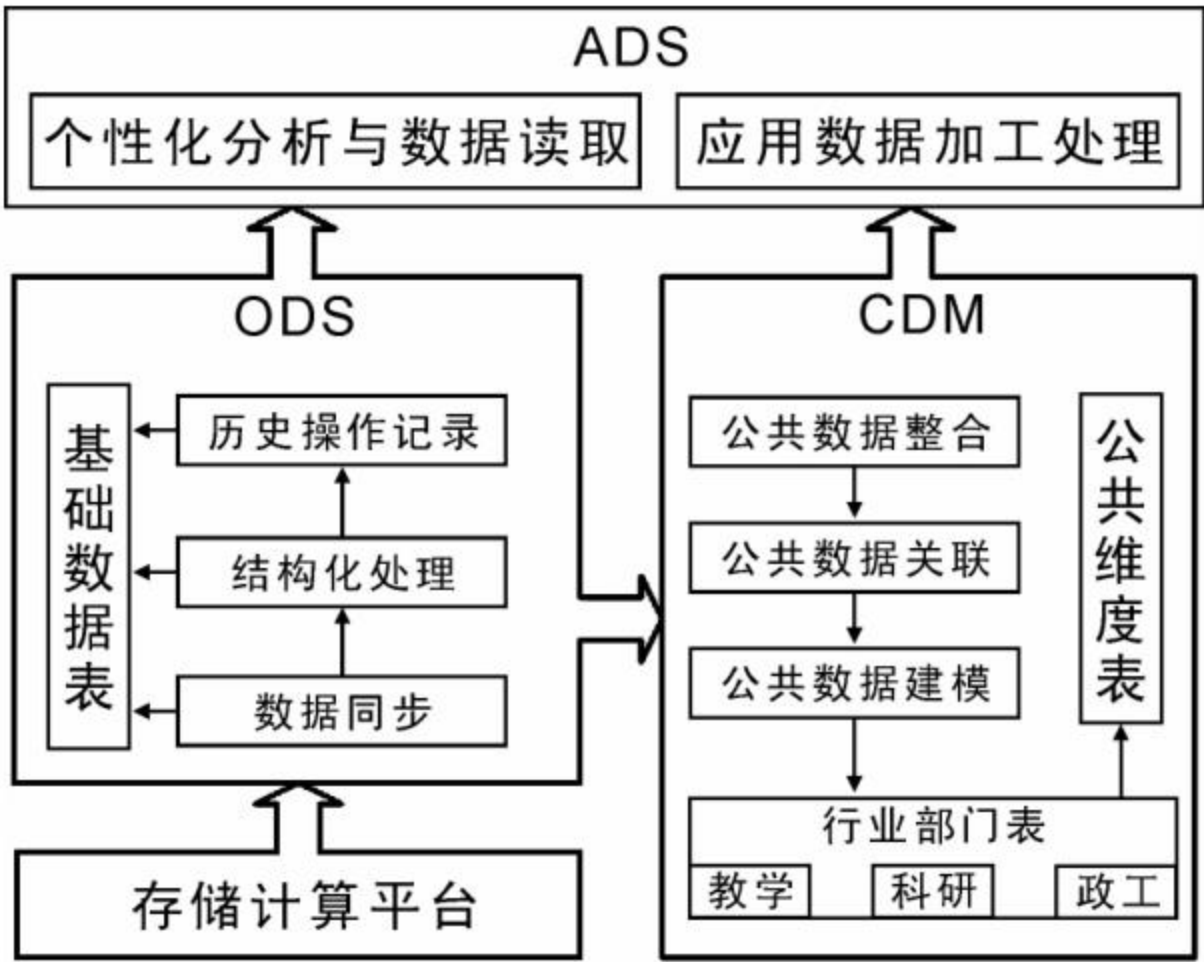


图 4 数据模型架构

ODS 层主要将存储计算平台中所采集到的各类数据进行数据同步。针对监控数据、门禁数据、教学业务系统数据和科研业务系统数据等进行数据同步和结构化处理,在此基础上构建基础数据表,同时记录历史数据的变迁和对底层数据的操作。

CDM 层主要完成基于维度的数据建模,针对不同业务模块和系统,进行数据的清洗过滤,划分和

定义数据域,构建公共维度表,为数据模型设计提供总体框架。例如,根据监控视频识别学员的上课坐姿和门禁系统的人脸识别情况等数据,得出某时间段内学员上课听讲情况或课堂受欢迎程度。

ADS 层主要功能是加工生产个性化的统计数据,同时提供数据读取的 RESTful 接口,为进一步的数据挖掘和数据开发提供基础。

1.1.3 数据开发

数据开发是数据资产内容重构的主阵地,是数据价值生成的核心环节。数据开发主要用于快速构建完整的数据挖掘分析工作流程,提供针对各类数据资源的常规统计分析和大数据分析服务。针对大批量数据的处理和挖掘需求,应提供一整套完整可靠的数据加工和管理工具,为校园数据资产管理部门提供数据治理支持,将汇聚的数据库存敏捷地转换为高价值的数据资产提供给各类校园业务系统使用。数据开发一般包括离线分析、实时处理、数据挖掘等多种服务功能。

1.1.4 数据资产管理

该模型提供统一的数据资产管理,帮助信息运维人员对数据资产进行全局把控、清晰查看、快速使用、规范管理和高效治理,以实现学校数据资产的安全统筹管理,提升数据资产价值和利用率。通过数据资产管理,使用者能迅速查找到自己关心、需要使用的数据资产,快速支撑新应用的开发。通过数据资产管理,可以获知数据资产的使用是否合理、数据的账目是否清楚等。数据资产管理一般包括数据标准管理、元数据管理、主数据管理、数据资产目录管理、数据质量管理,以及必要的权限管理等^[7]。

1.1.5 数据服务

数据服务是指各类数据应用的创建、测试和发布,目的是实现数据服务的敏捷响应,降低业务系统对数据的获取难度,提升校园用户的数据体验和工作效率,彰显数据资产的真正价值,这是数据中台的核心价值。数据服务为上层应用提供统一的数据服务,支持对数据服务进行统一管理和发布,能够快速生成数据服务,包括数据服务的托管,覆盖设计、开发、测试、发布、调用、查找、运维监测、安全管控等数据服务各个生命周期阶段。利用数据服务,各业务部门可以快速搭建以数据服务为核心的系统应用,运维部门则能够对各种业务系统的数据服务进行统一管理和灵活调用,快速将数据服务通过数据共享服务系统进行发布,实现数据的服务化输出。数据

服务一般包括服务目录管理、服务搭建管理、服务应用管理和常用的服务质量分析等。

1.2 基于中台的数据治理

开展数据调研,制定数据管理标准,建设数据仓

库,并采用数据中台对学校的各类数据进行采集交换、质量检测、数据修复及成果输出,以确保为学校建立一套标准统一、分类清晰、维度完整、内容准确、质量可信的数据资源^[8-9]。实施步骤如图 5 所示。

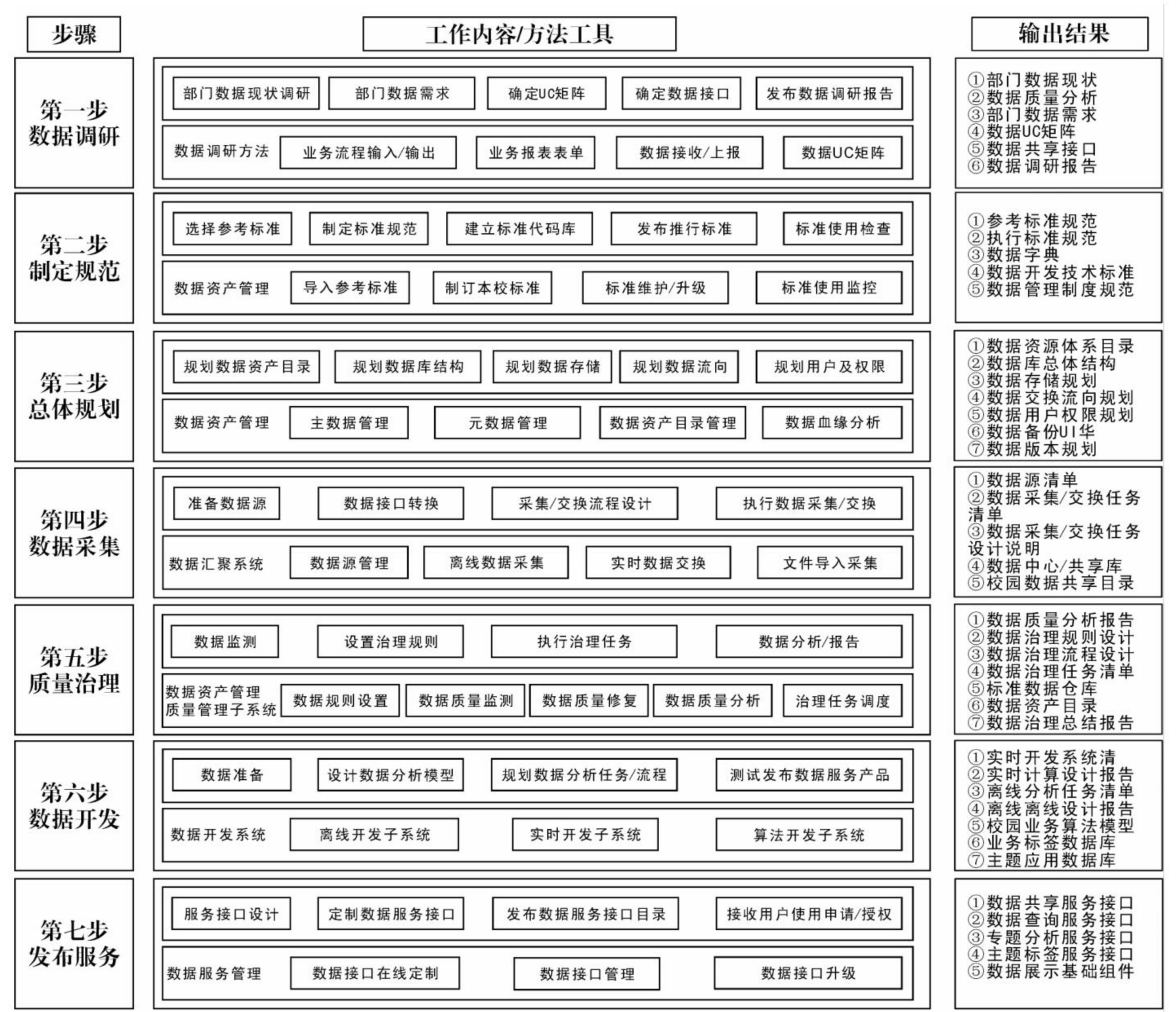


图 5 数据治理实施步骤

1.2.1 主要内容

数据治理工作一般可从管理、数据和知识三个方面开展具体工作：

①管理域主要是设计数据中台顶层架构、制定数据标准规范和完善数据管理制度机制。

②数据域主要包括数据工作、标准工作和质量管理工作三项内容。数据工作流主要包括数据辨识、数据采集、数据质量、数据存储、数据发布等。标准工作流主要包括设计元数据模型、编制数据标准、构建数据模型等。质量工作流主要包括数据质量检测、问题分析、数据转换、修正纠错等。

③知识域是指治理输出的数据域的结构性、知识性信息,主要包括:数据标准、数据源信息、元数据库、数据血缘关系、数据资产清单、数据 U/C 矩阵、数据交换关系、ETL 接口信息、数据质量规则库、数据治理检测报告、数据治理提升报告等。

1.2.2 方法步骤

数据治理的方法步骤如下：

①数据调研 对学校各部门组织架构、管理机制、管理信息系统、数据库现状进行深入调研,对数据的供需关系、共享交换要求、数据质量问题等进行深入了解,并输出数据现状报告。

②**制订标准** 依据国家标准、行业标准和其他垂直填报系统要求,结合学校业务现状,制定一套符合学校实际的数据标准(数据子集、标准代码集)。

③**总体规划** 通过数据调研,依据标准规范对全校数据建设、管理和使用过程进行顶层设计、总体规划,作为数据治理工作的指导原则。

④**采集交换** 通过数据汇聚组件对各业务系统的数据进行采集,存入共享数据库,等待清洗治理。

⑤**质量治理** 通过数据质量管理组件对采集上来的数据进行质量检测,并形成数据质量分析报告。对存在质量问题的数据进行修复、纠正,确保数据质量。

⑥**数据开发** 建立数据分析模型,通过离线开发或实时开发对标准数据仓库的数据进行分析计算,形成数据分析产品,支撑各业务领域数据综合分析需求,研发数据开发平台。

⑦**服务发布** 将开发好的数据集产品通过数据服务组件发布 API 接口,向各个应用系统提供数据调用服务。

1.3 体系架构特点

1.3.1 建成院校级数据标准规范

依据国家和行业智慧校园建设相关标准,依托数据中台技术对全校各部门、各业务中的数据需求进行全面调研和梳理,构建数据集合,划分数据子集分类,标定元数据内容,充分对接国家、行业标准代码和学校现有事实数据代码,结合学校实际情况,组织专业力量编制自有的代码标准、编码规范等。

1.3.2 建设统一的数据资产体系

数据资产体系是数据中台建设、管理、使用的核心要素,完成对全校统一数据标准的制定工作后,对全校各类数据资源进行全面的采集、治理,构建一套长期服务于全校的标准化、高质量、高价值的数据资源。按照纵向划分数据层级、横向划分主题域和业务过程的方式,构建一个结构清晰、层次分明、易于理解的数据资源体系。

1.3.3 构建完整的数据管理体系

制定数据管理相关规范和制度,涵盖数据资源的创建过程、数据中台的建设和日常维护以及数据服务的使用过程,包括数据管理的顶层设计、数据标准执行规范、数据访问安全规范、数据集成共享规范、数据质量保障规范、数据接口调用规范、数据安全保障要求等。

1.3.4 建设全面的数据服务体系

开展数据调研、数据采集、数据治理工作,通过数据中台提供数据共享、数据分析、数据可视化、数据查询服务,涵盖教学训练、科研学术、行政管理、后勤保障等学校各个业务领域,为智慧校园业务应用提供数据共享服务和大数据计算服务支撑。

2 实例分析

以海军航空大学智慧校园建设中的数据服务中心升级为案例,对采用数据中台技术升级架构前后的智慧校园数据各项情况进行统计分析,检验数据体系架构的高效性和可行性。主要针对采用中台架构前后数据对各项工作的支撑和影响为切入点展开,重点从数据利用情况、数据运维效率和数据服务支撑三个方面进行分析:

2.1 数据利用情况分析

数据利用情况主要针对智慧校园在利用数据中台技术升级后,用户活跃数、应用服务数量、数据吞吐量、IP 访问量等方面的信息展开统计分析。以 2022 年 1 月架构升级为分界点,统计之后的平均日用户活跃数、平均日 IO 吞吐量、平均日 IP 访问量等指标,与之前数据比较,如表 1(由于数据值本身涉密,设旧架构为基数 1 进行比较,2.2 和 2.3 节相同)所示。可以看出,在利用数据中台技术升级后,智慧校园整体服务对数据的利用情况明显活跃,日用户活跃数明显增加。

表 1 数据利用情况对比表

项 目	新架构	旧架构
用户活跃数(平均日)	1.634046053	1
IO 吞吐量(平均日)	3.795939273	1
IP 访问量(平均日)	1.261981344	1
应用服务总数量	2.153846154	1

2.2 数据运维效率分析

数据运维效率分析主要针对运维人员工作时间、运维账号在线时长、维修工单情况和厂家上门服务频次等指标数据,对智慧校园数据架构利用数据中台技术升级前后的情况进行对比分析,对考勤数据分析的同时,展开问卷调查,了解一线运维值班人员对中台技术加入后的直接反馈,如表 2 所示。从对比结果和运维人员反馈可以看出,利用数据中台技术升级架构后,运维人员在线时间有所下降,整体在工作时间之外加班时长也有所下降,体现了数据中台

技术对运维效率有所提高。另外,厂家技术人员上门服务次数有一定减少,节约了后期维护费用,提高了问题解决率,遇到故障平均修复时间有所降低。

表 2 数据运维效率对比表

项 目	新架构	旧架构
运维人员工作时间(平均日)	0.896 475	1
运维账号在线时长(平均日)	0.942 316	1
维修工单(次/年)	0.513 454	1
厂家上门服务频次(次/年)	0.744 156	1

2.3 数据服务支撑分析

通过统计架构升级之后,研究报告、科研课题和学习资料等教学科研数据共享的情况,来分析对数据服务支撑的效果。主要针对共享数据的下载次数,开设相关应用服务的咨询次数,问卷调查开展次数等指标展开统计分析,并对比升级前后数据,如表 3 所示。

表 3 数据服务支撑对比表

项 目	新架构	旧架构
共享数据下载(次/平均日)	1.42589	1
问卷调查开展(次/年)	3.26548	1
科研课题数据来源	1.75861	1
申请服务总数量	2.01544	1

通过对比分析可以看出,数据服务支持在利用数据中台技术升级架构之后有了一定程度的提升,申请服务和构建智慧应用的业务部门数量增多,同时教职员工利用数据挖掘技术开展教学科研活动、问卷调查等的数量也有了较为明显的变化,对科研工作 and 教学工作起到了很好的推动作用,也为数据中台建设带来了新的挑战。

3 结束语

数据中台逐渐成为当今各行各业数字化、智能化转型发展的基础,基于云计算、大数据、物联网、人工智能等新一代技术打造的企业级数据共享服务平

台,以丰富的数据计算和存储技术为数据中台提供了强大的数据处理能力^[10]。随着数据资源越来越丰富,数据使用场景越来越多元,数据中台将会融合更多新兴技术,释放出更多能量^[11]。

数据中台作为智慧校园的数据底座,对智慧校园的建设至关重要。文章提出的基于数据中台的智慧校园数据体系,可以实现智慧校园业务数据统一采集、数据资产统一管理、数据服务统一发布,为上层应用系统提供了数据查询、数据分析等服务,使数据资源在各部门间有效共享,真正发挥信息化效能,有利于助推学校转型升级,提高人才培养质量。

参考文献

[1]王智,尹长林,许文强.智慧城市背景下数据中台的研究与设计[J].网络安全和信息化,2021(7):29-35.

[2]杨敏,何海涛,张永强.基于数据中台的校园数据体系设计与建设[J].深圳大学学报(理工版),2020,37(S1):95-100.

[3]王雷,肖征,周晓明,郭辉.军队特勤疗养中心大数据中台的研究与设计[J].中国疗养医学,2021,30(11):1137-1139.

[4]张宏远.数据中台的通用体系架构研究[J].通信技术,2021,54(6):1451-1455.

[5]蒋东兴,付小龙,袁芳,等.高校智慧校园技术参考模型设计[J].中国电化教育,2016(9):108-114.

[6]张海华.基于数据轮询订阅模式的高校数据服务中台设计[J].信息系统工程,2021(2):30-31.

[7]崔恒志,王翀,吴健.基于数据中台的数据资产管理体系[J].计算机系统应用,2021,30(3):33-42.

[8]张翠翠,胡聪,洪德华,等.基于数据中台的数据安全研究与应用[J].现代计算机,2021(7):60-63.

[9]刘嫚,于磊.高校数据中心建设及数据治理探究与实践:以中国地质大学(北京)为例[J].中国教育信息化,2021(19):79-82.

[10]艾瑞咨询研究院.中国数据中台行业白皮书 2021 年[C].艾瑞咨询系列研究报告,2021(4):475-521.

[11]苏萌,贾喜顺,杜晓梦,等.数据中台技术相关进展及发展趋势[J].数据与计算发展前沿,2019,1(5):116-126.