

高职院校全量数据中心的關鍵实现路径

李斌

(三门峡职业技术学院,河南三门峡 472000)

[摘要]针对高职院校教育数据存在数据烟囱、质量不均等和融合服务度不足的原因进行了分析。其认为传统的共享式数据架构和数据仓库已经无法满足现阶段全量数据中心的建设要求,需通过数据交换作业设计将业务系统中的数据源加载到全量数据中心。高职院校全量数据中心建设的關鍵实现路径首先要健全数据标准体系,提升数据质量,其次自下而上依托于业务数据构建原始数据层 ODS、主题明细数据层 TDS、数据主题层 DWT 和数据应用层 DM 的四层架构实现高效率数据共享,并基于数据治理应用构建高职院校数据生态内外双循环体系,为数据驱动信息化服务打下坚实基础。

[关键词]数据分层;双循环;数据中心;高职院校;信息化

[中图分类号] G71; G434; TP311.13

[文献标识码] A

[文章编号] 2096-711X(2026)01-0065-03

doi:10.3969/j.issn.2096-711X.2026.01.023

[本刊网址] <http://www.hbxb.net>

引言

中国教育现代化 2035、教育数字化转型战略等国家政策和指导文件中对形成现代化的教育管理与监测体系,加快推进教育治理体系和治理能力现代化,推进管理精准化和决策科学化提出了明确要求,由此可见教育数据治理工作已经成为促进学校信息化发展的关键。在高职院校信息化领域,“创新驱动发展、数据驱动服务”已愈发成为智慧校园建设在面临新时代机遇和挑战下的基本战略。

面对以传统管理者为中心的多系统集中管理和底层烟囱式数据分布状态,如何转化为向基于角色的分散式应用系统和底层数据集中管理模式,重点是做到应用驱动向数据驱动转变,而数据驱动的前提是建设完善以盘活整体服务为中心,提升数据质量为重心,开放数据挖掘为核心的高职院校全量数据中心,促进高职院校各业务数据集成、数据交换、数据清洗、数据质量提升等数据治理工作,从而建立高质量的数据管理体系和统一的基础数据库,更好地支撑高职院校各业务系统高效运行,提升服务能力,促进对数据价值的挖掘和服务的再利用。针对全量数据中心建设的现实问题,本文提出高职院校全量数据中心建设的關鍵实现路径。

一、高职院校全量数据中心的现实问题

(一)数据烟囱

高职院校多年来建设的各类信息化系统积累了海量数据,师生的教育教学行为特征数据在各系统中得以采集和存储。但各系统建设时间差距较大,缺乏统一规划管理,数据标准设立滞后,造成各系统之间数据共享难度大,导致高职院校各机构和二级学院的数据难以实现衔接与共享,应用系统的“信息孤岛”造成了信息数据的“数据烟囱”。为打破数据烟囱,使各业务系统互联互通,即重点通过发掘盘点、整合集成、治理挖掘等步骤有效地提升数据治理能力,加速数据价值转换,从而实现数据同源,为高职院校的决策支持、教学科研、管理服务等多个方面提供强有力的数据支撑。

(二)数据质量

教育数字化转型产生了前所未有的多源异构数据,多源指的是数据来自不同源头,如传感器、埋点、应用系统、数据库、文档等,异构则指数据格式和类型不一致,包括结构化、半结构化和非结构化的数据。高职院校的教育数据又有自身的典型特点,表现为数据时间空间聚合差异度较大、数据源头关联度复杂,比如教育教学时间内外、关键考核评价点前后数据峰值差异明显、数据质量不均等。对不同数据源的异构数据进行综合分析和统一管理,提升数据来源质量,开发有效的数据采集处理方法尤为重要。对多源异构数据进行预处理首先要将多种异构数据源通过灵活的 ETL(提取、转换、加载)数据开发和任务引擎工具,实现数据有效接入共享处理平台。其次,需要使用数据开发中的节点和算子对数据进行处理。例如,数据清洗、数据合并、数据关联等操作对数据进行清洗和转换,以确保数据的质量和一致性,便于后续的分析处理。

(三)数据业务服务融合

随着业务系统的增加和使用量增加,高职院校应用业务系统产生的数据量越来越多。但因为缺乏一套整体规划的数据及服务体系,无法从根本出发在业务服务上做到更有效的数据资产化管理,随之而来的是大量冗余数据和垃圾数据的产生,使得数据的有效治理和高效应用无从下手,相关业务部门面对此种挑战存在畏难逃避心理,导致大量的业务流程阻滞。

传统的共享式数据架构和数据仓库已经不能满足日益增长的需求变化,需要以高职院校数据标准为依据,通过数据交换作业设计将业务系统中的数据源加载到全量数据中心,对高职院校整体业务服务、数据服务进行提升,梳理业务流程,构建服务中台,反哺数据中心,重构业务架构。

二、全量数据中心建设的实现路径

全量数据中心整体设计思路要遵循顶层统筹、以人为

收稿日期:2025-6-13

基金项目:本文系 2025 年度河南省高等学校重点科研项目“高职院校多源异构数据治理应用关键技术研究”阶段性研究成果(项目编号:25B520069);2023 年度三门峡市科技攻关计划项目“基于新型组合智能优化算法的卫星任务编排问题研究”阶段性研究成果(项目编号:2023L02003)。

作者简介:李斌(1974—),男,河南洛阳人,高级工程师,硕士,研究方向:教育信息化、数据分析。

本、数据智能、分步实施的原则,利用大数据、人工智能、物联网、移动互联网、5G 通讯、区块链等先进的 IT 技术,实现被动要求师生使用应用向信息化主动服务于师生转型。

(一)健全数据标准体系

前述数据烟囱和数据质量问题有着高职院校信息化发展的历史客观性,正视和解决问题的关键是信息化建设的体制机制创新,需要进一步明确不同层面、不同部门的权责,建立配套保障体系和制度。其中最主要的还是落实 CIO 工程,统筹调度分配资源,完善全量数据标准体系为后期数据打下基础。

首先,建立统一的数据标准,为中心库的构建以及数据的统一管理提供统一的标准参照,为数据的使用、数据共享、数据挖掘分析等打下基础。其次,构建全量数据接口标准,将所有的系统涉及的数据标准重新梳理,将公用的代码及相关字段统一,为全校的数据交换打下坚实基础。最后,建立数据 U/C(Use/Create)矩阵以及数据流图,定义数据使用消费的源头以及规范数据上下行,为数据交换共享和质量管理提供科学依据。

(二)四层数据架构体系建设

全量数据中心分层次建设有利于改善原数据中心表、视图混合混乱难以维护的情况,更好地为上层应用提供支持。自下而上依托于业务数据搭建四层架构的全量数据中心,通过基础业务主题数据四个层级架构及可视化的元数据管理功能,以数据采集过程为主线,推进数据服务体系建设,让数据来源于业务同时服务于业务,实现统一数据标准和高质量的数据共享,为高职院校提供更方便快速的全量数据中心管理。

1. 原始数据层(ODS)

ODS 层面向业务过程和事件,通过标准化接口直接从高职院校各业务系统和数据源中抽取、清洗和传输原始数据,通过建立数据拉链机制和质量监控体系,确保数据完整性和准确性,解决了传统数据接入方式中存在的接口不可控和数据质量难验证的问题。

2. 主题明细数据层(TDS)

TDS 层对高职院校核心业务数据进行统一管理和标准化处理,按照业务主题构建结构化数据集市,有助于数据的标准化存储。通过规范化数据组织和分类存储,既支持细粒度的主题数据查询,又实现了数据的可视化管控,为上层分析提供清晰的明细数据视图。

3. 数据主题层(DWT)

DWT 层面向高职院校数据分析需求构建聚合层,通过预定义的统计模型和计算规则,生成周期性、累积性多维度汇总指标数据。这些标准化统计结果既可作为数据质量核查基准,又能直接支撑分析应用,有效解决传统分析模型中存在的标准不统一和管理分散的问题,可显著提升数据分析效率。

4. 数据应用层(DM)

DM 层明确定义了数据服务接口的数据交互协议和访问控制策略,实现数据安全可控开放以生成统计报表。通过采用统一的接口规范管理数据服务,在保障数据安全的前提下,解决传统数据服务模式中存在的接口异构和管理混乱问题。

高职院校全量数据中心的四层架构通过标准化的数据流转和处理机制,实现高职院校从原始数据接入到服务开放的全链路治理,可以显著提升高职院校数据资产的管理效能

和使用价值。

(三)数据治理双循环体系

为实现对高职院校教育数据准确性、唯一性、权威性的需求,促进数据流向可控、数据使用可控、数据权威可控、数据安全可控,按照基于数据中台建设的数据治理应用和研究思路,构建高职院校高质量的数据生态循环体系,把数据生态划分为内循环和外循环两个生态循环。

内循环基于中台化的理论,构建出一个以数据中心为核心、以数据标准参照、以数据整合为服务、以质量检测为手段、以共享开放为目标的数据资产化管理中心,构建出有序健康的数据生态圈。内循环的关键点是全量数据中控台,它是统合管理全量数据中心各子平台及业务功能模块的统一操作平台,引入可视化数据地图的管理理念,通过数据流向图、数据血缘分析、数据转换关系可视化定义、历史数据拉链等功能的补充和完善,实现深度数据治理的可视化管控,进一步保证数据治理内循环的运行。全量数据治理内循环体系由数据标准管理系统、元数据及关系管理系统、主数据管理系统、数据清洗与整合平台、数据质量管理系统、数据资产管理系统和全量数据开放服务系统组成,包括全量数据中台的所有子平台对数据流转、数据质量检测、数据开放的个数、对象、总流出流入数据量、采集过程中整合的系统数量、文件数据源数量、检测规则数量进行同步监控和统计等。

数据治理外循环的要义可以总结为“一出一进,一少一多”。“一出”就是为内循环积累的数据找出路,不仅仅给其他业务系统提供数据接口,更要为使用者提供数据服务,让师生受益于数据治理的成果。“一进”是为了解决数据不足的问题,由于信息化进程的不平衡,有些数据在个人或部门的终端里没有进入业务系统,亦即在内循环过程中进入不到数据中心存在大量的数据缺失,后期只能通过手动录入的方式,既会增加人为工作量也会降低数据质量,外循环提供灵活的数据入口,方便各种终端数据进入数据中心,极大丰富数据中心的数据量。“一少”是减少业务部门的数据纠错负担,给师生提供数据服务的同时给予数据纠错的功能,使得使用者发现数据问题时能及时纠错。“一多”是提供更灵活多样的数据应用服务,除了提供个人数据服务和综合校情分析等定制化数据应用以外,还可以为师生提供需要的各种数据表,让师生自己创造一些基于数据的个性化应用,让高职院校师生有更佳获得感。

在数据治理中控台平台中,数据标准实现高职院校数据中心的完整性、有效性、一致性、规范性、开放性和共享性管理,为数据资产管理活动提供参考依据。元数据管理系统全量数据标准管理系统联动,真实体现采标率情况。同时管理全量元数据的关系及全量元数据的关系展现。全量主数据管理致力于高职院校实现主数据的标准化管理、统一数据表和全局共享,打通各信息系统之间的数据集成和业务协同。数据清洗整合系统进行不同数据结构间的数据转换,解决“脏数据”的清洗,完成异构多数据源的数据整合。数据资产管理全方位整理统计数据资产,让资产分类更简洁明了以避免资产目录分类混乱。数据质量管理有针对性地创建全面的数据质量业务规则库,同时结合数据整合清洗经验制定高职院校数据质量的量化指标,产生数据质量报告以利于整改。

(四)全量数据中心驱动数据服务

全量数据中心的核心价值观,体现于数据服务。优化数据

管理模式,实现高职院校全量数据的实时可采集、全面可管控、处处可复用,真正实现“数入一库,数出一库”,强化数据治理管理能力,提升数据质量,形成数据共享、数据治理的数据闭环。

基于全量数据中心,通过数据治理内外双循环体系建设,对高职院校数据进行中台化管理,形成良好的数据生态。将高职院校内外割裂的数据进行汇聚、治理、建模加工,为进一步实现受众者立体画像、智能数据推荐、业务实时监控等方面的应用提升数据驱动服务。

三、结语

高职院校全量数据中心建设要在健全数据标准体系和提升原始数据质量的前提下,重点搭建四层架构实现高质量数据共享,并基于数据治理应用构建高职院校数据生态内外双循环体系。全量数据中心建设的持续推进和数据资源的有效利用,在逐步消除数据烟囱将院校内各类人、财、物的公共业务及核心数据统一管理,彻底打通不同口径的数据与业务流程实现数据的重构与融合后,将为进一步构建统一身份认证、服务门户、网上办事大厅、一表通等智能化系统应用,整体提升高职院校的协同办公能力、工作效率和信息化服务能力,帮助管理者高效精准了解院校数据建设情况进而辅助决策打下坚实基础。

参考文献:

[1] 杨现民,李新. 数据充分赋能教育数字化转型的内在逻辑与实践路径[J]. 中国教育学报,2024(2):9-15.
[2] 吴岩. 深入实施教育数字化战略行动以教育数字化支撑引领中国教育现代化[J]. 中国高等教育,2023(2):5-10.

[3] 胡开开,李岩. 高校数据治理探究与实践[J]. 信息与电脑(理论版),2024,36(8):140-142.

[4] 曾敏,应巨林,郑乾生,等. 数据中台下的高校数据治理建设的应用与研究——以温州科技职业学院为例[J]. 办公自动化,2023,28(10):51-54.

[5] 田威. 职业院校数据中心建设探究与实践——以常德职业技术学院为例[J]. 科学咨询(科技·管理),2022(11):81-84.

[6] 王若竹. 基于数据生命周期的高校全量人员中心的建设与实践——以南京中医药大学为例[J]. 现代信息科技,2024,8(11):135-140.

[7] 王纯江. 可视化建仓与 ETL 工具的设计与实现[D]. 成都:电子科技大学,2024.

[8] 靳晓娟. 基于大数据技术的高职院校全量数据资源池的建设与实践[J]. 信息与电脑(理论版),2024,36(3):232-234.

[9] 银黔丰. 高校数据融合中数据标准与数据同步技术研究[D]. 长沙:湖南师范大学,2021.

[10] 孔晓宇. 基于大数据技术的数据中心建设与应用[J]. 无线互联科技,2021,18(7):98-99.

[11] 李引,陈敏锋. 高校数据中心生态的研究与应用[J]. 现代信息科技,2020,4(19):99-103+108.

[12] 苏舟. 某高校信息化基础平台设计与实现[D]. 成都:电子科技大学,2020.

[13] 陶行知. 生活即教育[A]. 陶行知全集(第一卷)[C]. 长沙:湖南教育出版社,1983.

Key Implementation Paths for Full-scale Data Centers in Higher Vocational Colleges

LI Bin

(Sanmenxia Polytechnic, Sanmenxia Henan 472000, China)

Abstract: Analyses were conducted on the reasons for the data chimney, quality inequality and insufficient converged service levels in education data in higher vocational colleges. It is believed that the traditional shared data architecture and data warehouse can no longer meet the current construction requirements of the full data center. The data source in the business system needs to be loaded into the full data center through data exchange operation design. The key implementation path for the construction of full-scale data centers in higher vocational colleges must first improve the data standard system and improve data quality. Secondly, from the bottom up, we build a four-layer architecture of ODS (operational data store), TDS (theme detail store), DWT (data warehouse theme) and DM (data mart) based on business data to achieve high-efficiency data sharing, and build a dual circulation system for data ecology of higher vocational colleges based on data governance applications, laying a solid foundation for data-driven information services.

Key words: data stratification; dual circulation; data center; higher vocational colleges; informatization

(责任编辑:桂彬彬)