

全国职业院校大数据中心建设指南

教育部教育管理信息中心

2023.06

目 录

第一章 总则	1
一、编制背景	1
（一）国家教育数字化战略行动	1
（二）职业院校数字校园建设	1
（三）国家智慧教育平台建设	1
二、编制目的	2
三、编制原则	2
四、编制单位与适用范围	2
五、主要参考标准与规范	3
六、指南发布网址	3
第二章 总体要求	5
一、指导思想	5
二、建设目的	5
（一）为职业教育人才培养创新提供服务	5
（二）为职业院校管理现代化提供服务	5
（三）为国家职业教育管理与决策服务	5
三、建设原则	5
（一）坚持目标导向、统筹推进	5
（二）坚持问题导向、有效实施	6
（三）坚持服务导向、应用为王	6
（四）坚持能力导向、创新发展	6
（五）坚持安全导向、协同共治	6
四、建设目标	6
（一）制定数据标准，完成数据治理，实现数据共享	6
（二）建立大数据底座，汇聚数据资源，支撑数字化转型	7
（三）挖掘数据价值，建设智能化应用，提供数据服务	7
（四）构建多级联动机制，实现平台对接，支撑职教智慧大脑	7
五、建设内容	7
（一）顶层设计	7
（二）数据标准与数据架构设计	7
（三）数据质量与安全管理	8
（四）数据应用	8
（五）大数据平台建设	8
（六）全国职业教育智慧大脑院校中台对接	8
第三章 顶层设计	9
一、数据战略	9
（一）数据战略规划	9
（二）数据战略实施	10

(三) 数据战略评估	11
二、数据治理体系	11
(一) 数据治理组织	11
(二) 数据治理制度	11
(三) 数据治理沟通机制	12
三、数据生存周期过程管理机制	12
(一) 数据需求阶段管理	13
(二) 数据设计与开发阶段管理	13
(三) 数据运维阶段管理	13
(四) 数据归档或销毁阶段管理	13
第四章 数据标准与数据架构设计	14
一、数据标准	14
(一) 数据元	14
(二) 参考数据	15
(三) 主数据	15
(四) 指标数据	16
二、数据架构设计	17
(一) 数据模型	17
(二) 数据分布	18
(三) 元数据管理	19
(四) 数据集成与共享	20
第五章 数据质量与安全管理	21
一、数据质量	21
(一) 数据质量需求和管理目标	21
(二) 数据质量检查	21
(三) 数据质量分析与提升	22
二、数据安全	22
(一) 数据安全策略制定	22
(二) 数据安全治理	23
(三) 数据安全审计	25
(四) 数据安全技术支持	25
第六章 数据应用	26
一、数据分析	26
(一) 对象分析	26
(二) 主题分析	27
(三) 综合分析	27
二、数据服务	28
(一) 党建引领	28
(二) 人才培养	29
(三) 学校治理	31
(四) 社会服务	33

第七章 大数据平台建设	34
一、支撑环境建设	34
(一) 存储计算资源	34
(二) 网络安全环境	34
二、平台技术框架	34
(一) 数据源层	35
(二) 数据汇聚层	35
(三) 数据存储与管理层	36
(四) 数据分析与开发层	37
(五) 数据应用层	38
(六) 数据治理监控层	39
三、建设模式	40
(一) 院校自主建设模式	40
(二) 政府主导的区域建设模式	40
(三) 校企合作建设模式	41
第八章 全国职业教育智慧大脑院校中台对接	42
一、数据对接内容	42
二、数据对接方式	42
三、数据对接技术规范	42
(一) 数据对接流程	42
(二) 数据接口规范	43
(三) 数据安全要求	43
四、数据对接准备	43
(一) 对接账号申请	43
(二) 数据整理	44
(三) 数据标准开发	44
(四) 数据接口集成	44
五、数据对接接口开发规范	44
(一) 接口调用方式	44
(二) 数据对接接口调用规范	45

第一章 总则

一、编制背景

（一）国家教育数字化战略行动

党的二十大报告指出“推进教育数字化，建设全民终身学习的学习型社会、学习型大国”。教育数字化是数字中国战略的重要组成部分，是推动教育强国建设的重要战略支撑。教育部《关于加强新时代教育管理信息化工作的通知》指出要利用新一代信息技术提升教育管理数字化、网络化、智能化水平，推动教育决策由经验驱动向数据驱动转变、教育管理由单向管理向协同治理转变、教育服务由被动响应向主动服务转变。国务院《“十四五”数字经济发展规划》提出“深入推进智慧教育”，强调推进教育新型基础设施建设，深入推进智慧教育示范区建设，推动“互联网+教育”持续健康发展。《教育部2022年工作要点》提出“实施教育数字化战略行动”，加快推进教育数字化转型与智能升级。数据体系建设是实现教育数字化战略的核心内容。

（二）职业院校数字校园建设

近年来，职业教育领域多个重要文件都将信息化与数据体系建设作为重点内容，《国家职业教育改革实施方案》要求“适应‘互联网+职业教育’发展需求，运用现代信息技术改进教学方式方法，推进虚拟工厂等网络学习空间建设和普遍应用。”《职业教育提质培优行动计划（2020—2023年）》指出需落实《职业院校数字校园规范》，推动各地研制校本数据中心建设指南，指导职业学校系统设计学校信息化整体解决方案。

（三）国家智慧教育平台建设

数字化赋能教育管理转型升级，是推进国家教育数字化战略行动的重要任务，是创新教育治理理念和方式的重要举措。教育部组织建设国家智慧教育公共服务平台，涵盖基础教育、职业教育和高等教育，建立教育数据大脑，加强数据挖掘分析，构建决策分析模型，探索基于新兴技术的新型教育政策研究模式，完善教育管理中枢决策系统的指挥调度功能，助力管理者科学决策，体现智能化教育管理的科学价值。

二、编制目的

在教育数字化快速发展的背景下，依据职业院校数据能力现状，结合职业院校大数据中心建设的主要实践编制本指南。通过本指南为职业院校、相关企业提供职业院校大数据中心建设、应用、运维的规范和指导，帮助职业教育工作者理解把握职业院校大数据中心建设中的问题和解决方案，应对职业院校数据管理挑战，助力职业院校治理水平提升。

三、编制原则

全国职业院校大数据中心建设指南（以下简称指南）按规范格式撰写。指南编制遵循以下原则：

1) 实用性与适用性

指南阐述职业院校大数据中心建设内容与方法，面向当前各类职业院校大数据中心建设中的实际问题和需求给出具有实用性和适用性、能够指导当前和未来一段时间职业院校大数据中心建设与管理工作的规范。

2) 科学性与可靠性

指南的编写以主流的数据管理框架为指导，参考国内现行的大数据和数据治理相关标准，并结合职业院校数字化战略的发展目标，给出科学、严谨、明确的规定，保证其具有参考的科学性和可靠性。

3) 准确性与可操作性

指南内容的语言力求简练、准确，不引起歧义；指南提供的体系框架在职业院校大数据中心建设中具有可操作性。

4) 共享性与先进性

指南的编写充分考虑数据互联互通需求，为与全国职业教育智慧大脑院校中台数据对接以及各级教育主管部门和跨院校数据共享提供参考。指南给出的建设方法使用当前大数据相关的主流技术与工具，保障大数据中心建设的先进性。

四、编制单位与适用范围

本指南由教育部职业教育与成人教育司指导，由清华大学、教育部教育技术与资源发展中心（中央电教馆）、教育部教育管理信息中心组织起草。

本指南主要编制人：韩锡斌、周潜、石凌、陈琳、刘英群、赵林、葛连升、沈富可、陈怀楚、宋继华、陈明选、张鹏、任怀德、杨小兵、李绍中、詹增荣、贾琳、陈晓东。

本指南适用于职业院校大数据中心的规划、建设、管理和应用，以及在全国职业教育智慧大脑院校中台的数据对接。

五、主要参考标准与规范

- [1]. 教育部.《职业院校数字校园规范》（教职成函〔2020〕3号）.2020-6-16.
- [2]. 教育部.《教育管理信息 教育管理基础代码》（JY/T 1001）.2012-3-15.
- [3]. 教育部.《教育管理信息 教育管理基础信息》（JY/T 1002）.2012-3-15.
- [4]. 教育部.《教育管理信息 中职学校管理信息》（JY/T 1005）.2012-3-15.
- [5]. 教育部.《教育管理信息 高等学校管理信息》（JY/T 1006）.2012-3-15.
- [6]. GB/T 36073-2018.《数据管理能力成熟度评估模型》（DCMM）.2018-3-15.
- [7]. 中国信通院.《大数据白皮书（2022年）》.2023-1-4.
- [8]. 中国通信标准化协会.《数据资产管理实践白皮书（5.0）》.2021-12-20.
- [9]. 国际数据管理协会.《DAMA 数据管理知识体系指南》.2020-5-1.
- [10]. GB/T 35295-2017.《信息技术 大数据 术语》
- [11]. GB/T 42450-2023.《信息技术 大数据 数据资源规划》
- [12]. GB/T 38672-2020.《信息技术 大数据 接口基本要求》
- [13]. GB/T 38675-2020.《信息技术 大数据 计算系统通用要求》
- [14]. GB/T 37721-2019.《信息技术 大数据分析系统功能要求》
- [15]. GB/T 37973-2019.《信息安全技术 大数据安全管理指南》
- [16]. GB/T 38633-2020.《信息技术 大数据 系统运维和管理功能要求》
- [17]. GB/T 34960.5-2018 《信息技术服务 治理 第5部分:数据治理规范》

六、指南发布网址

教育部教育管理信息中心门户网站（<http://www.emic.edu.cn/>）的“中心服务”区域开设“职业院校数字校园建设”专栏，提供《全国职业教育智慧大脑院校中

台高职数据标准及接口规范(试行)》、《全国职业教育智慧大脑院校中台中职数据标准及接口规范(试行)》、《全国职业院校大数据中心指南》的下载使用，职业院校数字校园建设试点工作进展和经验分享以及常见问题解答等服务。

第二章 总体要求

一、指导思想

以习近平新时代中国特色社会主义思想为指导，全面贯彻落实党的二十大精神，遵循创新、协调、绿色、开放、共享的新发展理念，充分利用大数据与人工智能技术，以职业教育信息化环境和数据资源为基础，以提升职业教育治理能力和人才培养质量为目标，构建并完善数据管理体系和技术平台。

二、建设目的

（一）为职业教育人才培养创新提供服务

职业院校大数据中心建设应以人才培养创新为首要任务和核心目标，结合职业教育特点和发展优势，用数字化思维重构职业教育生态，服务数字化时代的技术技能人才培养。

（二）为职业院校管理现代化提供服务

以职业院校信息化标杆校建设为契机，构建大数据中心，完善信息系统关键采集指标，实现对学生成长、教师发展、人才培养与校园管理等全流程数据实时采集与监测。充分发挥大数据中心的作用，提升数据应用的深度和效率，通过建立大数据分析模型推动教育科学决策和精准管理，有效支撑招生就业、学生管理、教学管理、校企合作、资产管理、安全管理等院校业务。

（三）为国家职业教育管理与决策服务

加强职业院校大数据中心规范化和标准化，通过跨地域、跨层级、跨部门的数据共享，支持职业教育管理部门全面、精准地掌握职业院校人才培养情况，为职业教育改革提供数据支撑。通过与全国职业教育智慧大脑院校中台对接实现职业教育数据的动态采集，支撑全国职业教育智慧大脑院校中台数据建设，及时客观地反映学校运行情况。

三、建设原则

（一）坚持目标导向、统筹推进

准确把握信息时代职业教育发展的重要特征，紧紧围绕立德树人根本任务，聚焦学校核心发展战略，做好顶层设计、制定数据战略规划。一体化布局，调

动全校师生和各部门的积极性，切实破解阻碍教学与管理数据共享开放的制度性瓶颈，统筹推进数据共建共治共享。

（二）坚持问题导向、有效实施

聚焦职业院校数据管理体系不健全、数据质量和安全难保证、数据管理能力成熟度不高、数据应用程度不深入等方面问题，充分借鉴 DCMM（Data Capability Maturity Model，数据管理能力成熟度模型）、DAMA（Data Management Association，数据管理协会）等主流数据管理框架与国标、部标、行标等相关标准规范，在本指南的基础上形成符合院校实际需求的数据管理规范 and 标准，保障大数据中心建设的有效实施。

（三）坚持服务导向、应用为王

从师生员工的需求出发，从面向用户的服务场景入手，以业务应用驱动数据治理，快速响应服务需求，提高服务质量。加强数据赋能，推进跨部门、跨系统协同与应用，扩展数据应用服务的广度和深度。

（四）坚持能力导向、创新发展

注重数据管理能力的提升，积极运用云计算、区块链、人工智能等新一代信息技术挖掘现有数据资源的价值，从而优化数据管理过程，提高数据质量，逐步达到较高的能力成熟度水平，支撑学校数据治理的创新发展。

（五）坚持安全导向、协同共治

树立数据安全底线思维，围绕数据生命周期，落实数据安全分级管理的主体责任，全面构建安全管理制度、管理机制和监测平台，形成协同共治的数据安全体系。

四、建设目标

（一）制定数据标准，完成数据治理，实现数据共享

推进学校数据标准建设，理顺数据规划设计思路，提供可行的数据支撑环境建设模式，以促进职业教育数据资源的共享与应用。开展数据治理，实现数据采集、清洗、存储、应用全生命周期的闭环管理，打破各部门的信息壁垒，实现数据共享。

（二）建立大数据底座，汇聚数据资源，支撑数字化转型

建设一套可靠的、先进的、可拓展的大数据平台，实现数据汇聚、存储、运算、开发、服务、可视化分析、质量安全管控一体化。以大数据平台为技术支撑，打造支撑学校数字化转型的大数据底座，服务学校教学和管理数字化，助力学校数字化转型。

（三）挖掘数据价值，建设智能化应用，提供数据服务

构建数据仓库，提供数据综合查询、主题分析、对比分析等数据服务，满足校内和校外各类人员的数据需求。构建智能决策支持系统，基于数据挖掘技术探索教育教学规律，满足学校管理需求。

促进师生不断使用各种形态的数据应用，实现“差异化的教、个性化的学、精准化的管、智能化的评”，提升学习、教学、管理的效率和质量。

（四）构建多级联动机制，实现平台对接，支撑职教智慧大脑

推进职业院校与国家智慧教育公共服务平台的对接工作，不断完善数据标准和对接接口，支持校本建设和区域建设两种大数据中心建设模式与平台的对接，逐步形成部、省、市、县、校多级数据联动机制。

五、建设内容

学校大数据中心建设主要包括顶层设计、数据标准与数据架构设计、数据质量与安全治理、数据应用、大数据平台建设、全国职业教育智慧大脑院校中台对接等内容。

（一）顶层设计

主要阐述保障大数据中心数据管理有效开展的主要工作内容，包括数据战略、数据治理机制、数据生存周期过程管理机制等。

（二）数据标准与数据架构设计

主要阐述制订大数据中心数据标准，及针对数据需求对数据进行分布梳理、模型重构的标准化过程。数据标准保障数据的内外部使用和交换的一致性和准确性，主要工作内容包括数据元标准、参考数据、主数据、指标数据等。数据架构主要阐述汇聚数据，发挥数据价值的过程，工作内容包括数据模型、数据分布、元数据管理、数据集成共享等。

（三）数据质量与安全管理

主要阐述保障大数据中心数据质量和数据安全的主要工作内容。数据质量包括数据质量需求和管理目标、数据质量检查、数据质量分析和提升。数据安全包括数据安全审计、数据安全管理和数据安全策略。

（四）数据应用

主要阐述数据服务与应用的数据分析方法和应用场景，数据分析方法包括对象分析、主题分析和综合分析，应用场景包括党建引领、人才培养、治理体系、社会服务等方面。

（五）大数据平台建设

主要阐述大数据平台的支撑环境、工具和平台建设模式。支撑环境包括网络基础设施、云计算基础设施，大数据存储与计算平台。工具包括数据标准与架构管理工具（含元数据管理、参考数据与主数据、指标数据、数据模型、数据集成与共享）、数据质量与安全工具、数据服务与应用工具（含数据开发工具、数据服务工具、可视化工具）等。建设模式包括（校本）建设模式、区域建设模式和校企合作建设模式，并就职业院校大数据中心构建提出建议。

（六）全国职业教育智慧大脑院校中台对接

主要阐述学校大数据中心与全国职业教育智慧大脑院校中台的数据对接内容、对接方式、对接技术规范、对接准备及对接接口开发规范。

第三章 顶层设计

数据是与土地、劳动力、资本、技术相并列的战略资产和生产要素，学校应充分挖掘数据资产的战略价值以支持学校实现战略目标，提升学校治理能力和核心竞争力。学校应在充分了解自身数据现状和需求的基础上，建设数据管理体制和决策机制，依照数据生命周期确定数据管理流程，建立数据管理组织机构和权责范围。学校应从整体战略出发进行顶层设计，确定数据战略目标，将目标逐层分解形成战略规划任务，最终确定分阶段的实施路径和行动计划。

顶层设计可分为数据战略、数据治理体系、数据生存周期过程管理机制等方面。

一、数据战略

数据战略是学校所有利益相关者之间就数据管理愿景和目标达成共识的结果，它从宏观和微观两个层面指导学校开展数据管理与应用。制定并持续维护数据战略规划，包括战略目标、范围和内容、实施策略、实施路径等，监督整个战略的实施并进行评估。

（一）数据战略规划

分析数据需求和现状。数据需求主要是识别利益相关者，明确相关者的需求。利益相关者包括数据生产者和数据消费者两类，在职业院校，学生、教师、各业务管理部门等既是主要的数据生产者，也是数据消费者，除此之外，上级管理部门、学校管理机构、合作企业等也是主要的数据消费者。不同利益相关者的需求反映了各自的愿景和目标，学校需要分析、梳理并协调这些需求，保证需求的内部一致性。此外需确定各类需求的优先级，以此作为制定数据战略短期、中期和长期目标的依据。数据需求主要来源于两个方面，一是业务层面，二是技术层面。业务层面的需求与数据的应用场景相关，包括人才培养、管理决策、业务协同等，技术层面的需求与影响数据价值发挥的技术因素相关，包括数据质量、数据安全、数据模型、数据集成和共享等。学校应盘点已有的业务系统和数据资产目录，依据 DCMM、DAMA 等数据管理和评估模型客观全面地评估现有数据能力，从需求和能力两个方面找出自身差距，准确定位学校当前的数据管理所处的阶段和水平。

制定数据战略愿景和目标。愿景和目标是对数据在学校战略中核心价值的

明晰表达，愿景体现学校的中长期战略，目标体现短期战略。学校应遵照国家对职业教育数字化转型的政策要求，依据当前的发展阶段和能力水平，参考同类院校的成熟做法和先进经验，针对不同利益相关者的实际需求和优先级顺序，划分发展阶段，确定每个阶段的目标和指标性效果检验要求。学校的愿景和目标受国家政策、产业发展、技术进步等多方面因素的影响，因此应合理制定短期、中期和长期目标以提高战略规划灵活性。学校的短期目标一般体现最基础、最迫切的痛点问题，如数据不标准、不一致、不共享等，可通过数据治理高效解决。中期目标实现学校教学改革和教育管理的创新，通过数据的精细化使用构建教育新业态、探索教育新模式等；长期目标可依据学校特色来制定，以帮助学校在教育竞争生态中明确角色、提高地位。

制定数据战略行动计划和实施策略。为了有效达成每个阶段的目标，学校应根据自身数据能力现状和资源配置情况，制定数据战略的行动计划和实施策略。行动计划是落实战略目标采取的任务分解、资源调度和进度安排，应具备可执行性、协调性和可监控性。数据战略的实施涉及学校多个部门，往往直接影响数万名师生，实施周期长，过程较为复杂，因此需要不断迭代、持续优化的实施策略，以确保取得预期效果。

（二）数据战略实施

建立数据职能框架。确定每个阶段的具体实施任务，明确任务目标和目标完成的评估准则。明确相关机构和人员在任务中的角色和职责，形成内部沟通协调的工作机制，确保学校数据战略的实施。

建立保障方案。依据行动计划和实施策略，确定各项实施任务所需的基础条件和资源，制订详细的预算方案，保障人员、资金和设备的投入。制定实施流程和规范，评估和跟踪各项任务的实施进度情况，形成工作报告制度，并结合工作进展调整更新行动计划，保障数据战略的稳妥推进。

培养数据文化。在数据战略落地的过程中，学校要构建数据文化，从意识上提升学校的数据能力。学校应通过定期的培训和宣传来提高师生的数字化素养、培养师生的数据觉察力、分析力和敏感度，使数据战略的落地实施事半功倍。

（三）数据战略评估

战略绩效评估。在学校数据战略实施过程中，对照战略目标和实施过程，从投入、产出、时间和保障支撑等维度对学校数据战略落地的内容和关键的指标开展战略实施评估。

数据能力评估。学校的数据能力评估应以 DCMM 框架为基础，融入 DAMA 数据资产管理等关键内容和学校及上级管理部门对数据服务的关键要求，形成了符合自身的数据能力评估框架。

二、数据治理体系

为维持学校数据建设、管理和应用的规范高效运行，学校应建立并不断完善数据治理体系，包括建立相应的组织、制度和沟通机制，并视自身情况配套数据治理相关系统，加强对数据流程、数据源、数据质量的管理，通过数据治理推动学校全要素、全业务、全领域和全流程的数字化转型。

（一）数据治理组织

数据治理组织是各项数据职能工作开展的基础，包括组织架构、岗位设置、团队建设、数据责任等内容。学校应梳理并建立明确的数据治理组织体系，包括全校数据治理相关的组织架构，岗位设置、责任权力、分工协作、人员培训机制，明确相关的工作内容和责任以及相应的工作绩效评价政策和机制。

组织的构成一般可以由数据管理委员会、数据管理执行部门构成。管理委员会是学校数据管理的最高决策机构，主要职责是审议学校数据管理整体方针、目标、策略、体系规划、政策制度；统筹协调资源解决数据管理领域重大事项；听取数据管理执行部门的工作汇报并对数据管理工作进行监督评价。数据管理执行部门可以是一个科室如信息中心数据管理科，也可以是一个处室如大数据赋能中心等。数据管理执行部门负责全校数据管理工作的具体执行，以及协调各个部门的数据管理人员。具体的工作包括了制定和完善数据管理工作规章制度和流程；开展数据管理相关工作并监督落实；指导推进各部门数据管理工作；整理各部门的数据需求并进行数据开发与反馈；开展数据应用的培训；提高师生数字化素养。

（二）数据治理制度

为保障数据管理和数据应用各项职能的规范化运行，学校应建立并不断完

善相应的制度体系，包括数据管理制度、工作流程等，覆盖数据采集、存储、流转、使用的各个环节，以及数据治理的各组织和工作岗位。学校数据治理制度应符合外部合规、监管要求，应根据实际情况不断修订和完善，定期开展相关培训和宣传，持续优化管理过程。

数据治理制度建设要围绕数据管理工作划分多个不同的层级，包括总体规定，管理办法，实施细则和操作规范。总体规定文件从决策层和组织协调层视角出发，阐述数据管理的目标、组织、责任等。管理办法是从数据资产管理层视角出发，规定数据资产管理各活动的管理目标、管理原则、管理流程、监督考核、评估优化等，如数据模型管理办法、数据标准管理办法等。实施细则是从数据管理层和执行层的视角出发，围绕管理办法相关要求，明确各项活动职能执行落实的标准、规范、流程等，如主数据实施细则、数据开发实施细则等。操作规范是从数据管理执行层的视角出发，依据实施细则，进一步明确各项工作需遵循的工作规程、操作手册或模板类文件等，如数据开发操作规范、数据管理平台用户手册等。整体制度体系架构可以参考《数据资产管理实战白皮书》。

（三）数据治理沟通机制

学校数据治理体系的构建应明确数据是学校的重要资产，大数据中心建设是全校一盘棋。应着力解决数据建设、管理和使用中部门机构之间相关责任不落实、分工协同不顺畅的问题。为此学校应建立沟通机制，建立沟通的路径和渠道来获取与分析各利益相关者的需求；制定定期沟通计划，促进各部门、全校师生对数据治理达成共识。制定培训计划并开展培训，提高师生数据应用能力。

学校数据治理应坚持“数据以共享为根本、不共享为例外”的原则，推动数据的开放共享。科学设计数据体系架构，充分发挥数据中心能力，创新数据应用，赋能职业教育科学决策、精准管理和师生服务。

三、数据生存周期过程管理机制

学校在数据规划、建设、管理、应用和服务过程中应实现数据的生存周期管理。数据生存周期管理包括从数据需求、数据设计和开发、数据运维、到数据归档或销毁的全过程，数据安全和数据质量管理贯穿其中。

（一）数据需求阶段管理

学校应建立数据需求的管理组织、制度和流程，能够基于数据应用场景收集相关数据需求并进行梳理和定义，对数据需求进行评审，实现数据需求统一管理。对于已有数据管理标准中尚未覆盖的数据需求和经过评审后达成一致的需求，则根据相关的流程更新数据管理标准。

（二）数据设计与开发阶段管理

学校应有规范的数据设计和开发流程，有解决方案的编制规范、质量标准和安全标准，能够设计、实施和维护满足数据需求的数据模型和解决方案，能够确保解决方案与数据架构、数据标准、数据质量的一致性，以及数据的完整性、安全性、可用性和可维护性。数据设计和开发有持续改进措施，能支撑数据战略的落地实施。

（三）数据运维阶段管理

制定能与数据架构、数据标准、数据质量等一致的数据运维方案，部署支撑实现数据供应方管理、数据需求管理和数据需求的变更管理等功能的数据运维工具和平台，能够衡量数据运维方案有效性，明确数据运维服务质量持续改进的措施。

（四）数据归档或销毁阶段管理

构建数据归档、数据迁移、数据销毁的相关制度、标准和流程，建立对历史数据的使用、保留和清除的方案，符合内外部业务需求和监管要求。

第四章 数据标准与数据架构设计

一、数据标准

学校建立数据标准，在全校范围内统一定义数据的含义、数值范围、分类、统计口径等，确保数据的完整性、有效性、一致性、唯一性、准确性，是实现数据集成共享、提升数据质量、实现数据服务和应用的关键。数据标准规范应明确每个字段的生产部门、系统，确保形成“一数一源”的数据管理规范。

学校建立数据标准，应符合或参考国家、教育部、行业标准，包括教育部《职业院校数字校园规范》，教育部《教育管理信息 教育管理基础信息》、《教育管理信息 高等学校管理信息》、《教育管理信息 中职学校管理信息》、《教育管理信息 教育管理基础代码》等。结合学校业务实际，采用制度约束、过程管控、技术工具等手段，制定和发布由数据关联部门确认的数据标准，进一步提升数据质量。建立数据标准制定、发布、管理、应用和宣传等方面的制度和流程，并定期对数据标准检查和修订。。

数据标准是数据管理的基础，为学校数据仓库建设提供规范化、标准化的依据。数据标准主要包括数据元、参考数据、主数据、指标数据等四部分。

（一）数据元

数据元作为数据标准的基本单元，它是由定义、标识、表示和允许值等属性描述的数据单元。学校应建立统一的数据元目录和数据元管理规范，通过制定数据元标准，使得数据拥有者和使用者对数据有一致的理解。

数据元规范制定。学校根据实际业务特征制定数据元规范，规范包含数据元的业务属性规范（如命名、版本等标识类数据和状态等）、管理属性规范（如提交机构等）、以及技术属性规范（如数据类型、数据格式、计量单位、值空间等）。

数据元目录构建。学校基于实际数据特征定义数据元分类规则，并根据分类规则和实际业务管理需求建立数据元的管理目录，数据元的分类可引用数据模型的主题域建设的分类，以《教育管理信息 高等学校管理信息》的数据集分类为参考，根据学校实际进行分类规则制定，如可以将一级分类定义为学校概况、教职工管理、学生管理、教学管理、科研管理、财务管理、资产管理、党建思政管理、办公管理、服务管理等。

数据元的管理。学校应制定数据元的管理流程，实现对数据元及目录的变更管理，并支持对数据元的分类查找与引用分析。

（二）参考数据

参考数据，也称数据字典，是用于将其他数据分类的数据，例如地区编码、专业代码、课程性质、性别、学历等。参考数据采用数据集合的形式规范并约束数据的值域，可被一个或多个枚举型数据元引用。学校应明确参考数据的引用来源和标准，并建立参考数据的管理规范和数值域的管理，包括分类的规范、标准化术语、代码值和其他唯一标识符。

参考数据描述。参考数据的核心是码值列表，包含代码与代码描述信息，必要时可添加说明性字段进行补充。当参考数据发生变更时，需要同时保留不同版本的标准，采用原始字典和标准字典加以区分，并建立新旧标准之间的映射，以便在数据处理时完成数据的转换。

引用标准设计。学校应在教育部《教育管理信息 教育管理基础代码》和《全国职业教育智慧大脑院校中台》中的编码基础上，参考国标、行标等通用标准体系，根据学校实际业务发展情况增加自定义的数据字典，不断健全完善参考数据标准编码体系。

参考数据分类。学校设计参考数据一级分类时可根据引用来源进行分类，可将《全国职业教育智慧大脑院校中台》作为一级分类，参考引用来源的分类建立二级分类。

（三）主数据

主数据是用来描述学校核心业务实体的数据，是跨越学校各个业务部门和系统的高价值的基础数据。主数据的应用范围最广，数据管理应重点保障主数据的稳定性、实时性、准确性和一致性。学校应明确主数据的来源、建立主数据的管理规范、做好主数据的准确记录，实现主数据跨系统的共享。

主数据治理。梳理现有业务系统主数据整合需求，识别可信数据源，明确主数据整合架构，按照学校业务规则定义主数据的数据模型和数据标准，明确数据质量规则和数据安全要求，保证数据的完整性、唯一性和合规性。定义主数据标准的管理机制，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，形成符合学校实际的管理流程、管理规范 and 制度、管理规则和考核机制。

主数据开发共享。明确主数据的业务范围、定义主数据的共享机制，建设和维护主数据应用仓库，依托支撑工具和平台实现业务系统和主数据存储系统的同步共享，保证跨系统的数据一致性和共享性。

（四）指标数据

指标数据是学校业务开展过程中用于衡量某一个目标或事件的数据，例如学生人数、专任教师人数、生师比等。指标数据管理指的是对学校开展各项教学与管理分析所需要的指标数据进行统一的规范定义、开发和应用，从而支撑学校师生的有效应用。学校应建立符合学校需求且满足上级管理要求的指标体系，建立统一的指标数据管理流程、明确指标数据的分类规范和格式规范、建立指标数据标准并清晰定义指标数据的含义、制定学校的指标数据分类并进行管理。此外，根据学校师生的需求，进行数据的访问授权，并对指标数据的应用进行监控，保证数据的准确性和及时性。

指标体系构建。构建包含原子指标、派生指标和复合指标等指标类型的指标体系：原子指标是反映教学或管理的某一个指标的最小表达，如成绩；原子指标+修饰词+时间周期构成派生指标；原子指标和派生指标之间的加减乘除运算可形成复合指标。修饰词是对指标进行的抽象业务限定，从而规定了指标所适用的所属修饰词类型的范围，比如学生性别类型，包含男性、女性等。此外，原子指标进一步细分为主原子指标和衍生原子指标，其中衍生原子指标需要由主原子指标+衍生词构成。衍生词是对于原子指标中带有计算口径的词进行的抽象定义，如成绩优异：平均成绩 90 分以上的学生，成绩优异为衍生词。

指标数据管理。指标数据管理定义各指标的创建、发布、变更等流程，每一个指标数据应明确指标的名称、标识、业务口径和业务口径负责人，可以根据需要增加技术口径，并明确负责人。

指标分类设计。指标分类（指标域）的管理是把比较接近的业务过程或者属性相近的内容划分为一个大的整体，作为对象分析和主题分析的基础，例如将指标域划分为学校概况、招生情况、师资发展、学生培养、教育教学、产教融合、实验实训、实习情况、党建工作、科技服务、毕业就业、服务水平等。

二、数据架构设计

学校应全面分析学校业务发展方向，理解把握学校的数据需求，明确业务数据、流程、组织和系统之间的关系，建立合理的数据架构，来保证数据能够高效地存储、集成、共享和使用。学校数据架构管理是通过对学校数据模型、数据分布、数据流转的管理，在全校范围内提供一个促进数据资产存储、集成、使用、访问和传输的框架结构组件，从而支持数据的集中管理和分析应用。

学校大数据中心的数据架构是针对学校多源异构的源数据场景，在数据组织层面为数据的接入、融合及应用服务等提供稳定、高效的支撑。因此，学校应制定数据架构规划，定义和维护数据架构原则与规范，开展数据架构的管理，开发和维护数据模型，明确可信数据源与数据流，以及数据分布与存储情况，审核系统设计是否符合数据架构整体设计等。

（一）数据模型

数据模型是将学校各项用于教学和管理决策的数据需求进行分析后，遵照数据模型设计规范使用结构化语言对需求重新组织和描述。学校应从数据视角对学校各项人、事、物的特征进行模拟和抽象，借鉴业界已有的数据模型相关方法和经验，编制学校数据模型设计规范，用于指导数据模型的设计和建设。

考虑学校数据的复杂性、多变性和安全性需求，建立与业务需求、组织结构以及管理流程匹配的数据模型，便于数据的采集、管理、流转和使用，而数据从接入、存储、开发、输出到展示，通常会构建数据仓库来贯穿整个流程。数据建模是数据仓库建设的第一步，良好的数据模型可以帮助学校更好地存储数据，更有效率地获取数据，保证数据间的一致性。

建模方法。可采用维度建模方法建设数据模型。维度建模方法以分析决策的需求出发构建模型，它重点解决用户如何更快速地完成需求分析，同时具备较好的大规模复杂查询响应性能，符合存放海量数据的数据仓库模型设计。维度模型一般采用星型模型，呈现的状态一般为星座模型。

模型架构设计。数据模型架构中将数据存储分为三层（如图 1 所示），包括数据贴源层（ODS）、数据仓库层（DW）和数据应用层（ADS）。数据仓库层又分为数据明细层（DWD）、维表层（DIM）、数据中间层（DWM）和数据服务层（DWS）。ODS 层主要负责存储从不同业务系统等来源汇聚过来的原始数据，数

据保持原貌不做处理。DWD层以业务过程为驱动，对ODS层的数据进行清洗（去除空值，脏数据，超过极限范围的数据）、维度退化、脱敏等，并基于维度建模，明细宽表，复用关联计算，减少数据扫描。DWS层以DWD为基础，进行轻度汇总，构建命名规范、口径一致的统计指标，为上层提供公共指标，建立汇总宽表。ADS层以DWS、DWD、DIM为基础，按主题进行汇总，为各种统计报表提供个性化数据、宽表集市、趋势指标。



图1 数据模型架构中数据存储的层次

主题域构建。主题域主要是面向业务过程，将业务活动事件进行抽象的集合，如教学管理、科研管理都是业务过程，主题域针对数据明细层（DWD）和维表层（DIM）进行主题划分。主题域的分类构建应以《教育管理信息 高等学校管理信息》的数据集分类为参考，并根据学校信息化的发展实际进行制定。如可以将一级分类定义为学校概况、教职工管理、学生管理、教学管理、科研管理、财务管理、资产管理、党建思政管理、办公管理、服务管理等。

指标域构建。指标域主要是面向业务分析，将业务过程或者维度进行抽象的集合。指标域是针对数据服务层（DWS）进行数据域划分，以上级管理要求和业务发展需求为驱动，划分方式可与指标分类保持一致。

（二）数据分布

针对数据模型中的数据定义，明确学校数据在系统、组织和流程等方面的分布情况，定义数据的来源及在各流程和系统间的流转情况，包含信息流、数据流、数据源等信息。通过数据分布的情况定义数据的分类方式、明确数据的

负责人、优化数据的集成关系。

数据源定义。数据分布的核心是数据源，数据源应在数据分布与数据标准中有明确的定义，并保证一数一源。学校梳理各个业务系统的数据源，定义标准数据的权威数据源，对每类主题数据明确相对合理的唯一信息采集和存储系统。

数据流设计。根据数据模型，学校应对数据进行合理的分类，建立数据分类管理机制，明确各类数据在系统中的关系。图 2 给出了在数据仓库的数据模型架构下数据分类和数据流关系图。

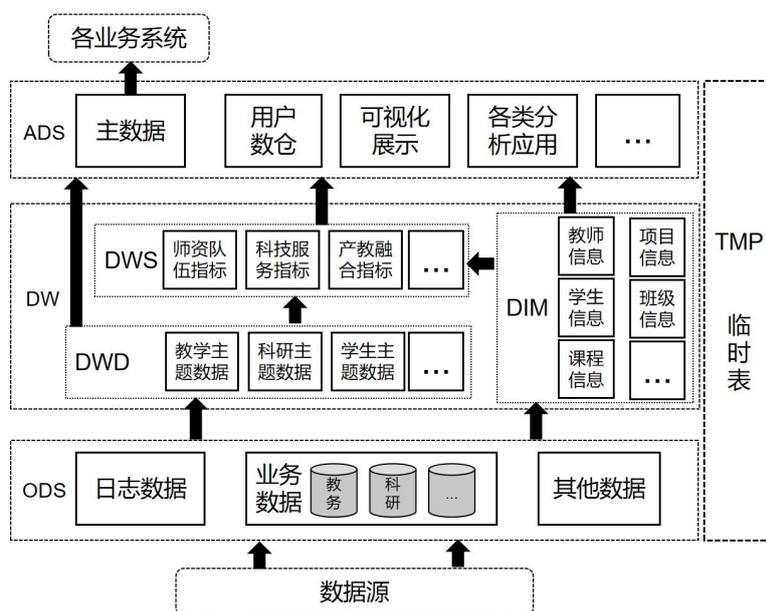


图 2 数据仓库的数据模型架构

（三）元数据管理

元数据管理是关于元数据的创建、存储、整合与控制等一整套流程集合。学校应根据业务实际过程、管理需要、应用的需求情况对元数据进行规范化管理，实现不同来源元数据的有效集成，提供元数据服务辅助数据管理与应用。

元模型管理。元模型用来描述元数据的属性定义。学校可采用和参考相关的国家标准定义各类元模型，建立元模型标准，规范元模型变更。

元数据集成和变更。基于元模型对不同类型和不同来源的元数据进行有效集成，形成数据全景图，并遵循相关规范及时更新和管理元数据变更。将元数据分为业务元数据、技术元数据、操作元数据和管理元数据，并从这四个维度管理和使用数据。

元数据应用与服务。基于数据管理和应用的需求，对各类型的元数据进行分析，提供元数据的查询、血缘分析、影响分析、符合性分析、质量分析等应用，提升全校师生对数据的理解，辅助数据管理与应用。

（四）数据集成与共享

数据集成与共享实现学校的各类数据在不同系统之间的互联互通。学校应建立各应用系统与大数据中心的数据交换规范和机制，构建数据集成共享平台保证各类数据在系统中高效、灵活地进行共享。集成平台需实现结构化和非结构化数据的处理、具备数据加工、挖掘分析和便捷访问功能。数据开放共享是实现数据跨部门、跨业务的重要前提，也是发挥数据价值的基础。学校应将内外部汇聚到大数据中心的数据按照统一的管理策略对全校师生进行有选择的开放。学校应该在安全、受监管的前提下促进数据的互通和数据价值的提升。

共享制度与标准制定。根据学校具有的数据资产，制定数据集成共享的原则、方式、方法，并根据共享方式制定数据的交换标准。

共享环境建设。建设能对结构化、非结构化等复杂数据进行加工处理、挖掘分析、服务共享的平台，并能依据数据交换服务规范，利用平台整合内部多种类型数据实现统一采集、集中管理与共享，实现大数据中心与各业务系统和上级应用平台的数据共享。

共享策略制定。学校应建立数据使用政策和开放共享策略，明确数据开放共享的质量、安全、流程要求。

数据资源目录编制。学校应在对数据进行全面梳理的前提下，形成本校的数据服务目录，按照一定的分类方法，对学校信息资源进行排序、编码、描述，便于检索、定位与获取信息资源。数据服务目录还需明确数据提供方、数据项、共享服务方式、提供方式、更新周期等内容清单，明确各部门数据共享责任，旨在推动建立学校数据共享机制。

共享个性化建设。学校可根据不同用户的角色、岗位、用户属性形成师生个人数仓、各部门各业务的数仓，来满足每一个师生个性化的需求，从而实现数据在学校内部充分的共享。

第五章 数据质量与安全管理

一、数据质量

数据质量是开展数据应用和服务的基础，是体现数据价值的前提。数据质量可以从完整性、及时性、准确性、一致性、唯一性、有效性六个维度进行描述。学校应围绕这六个维度，根据数据战略规划目标，从数据的使用者角度出发，明确各类数据的质量管理需求，制定学校数据质量管理目标、定期开展数据质量检查和质量分析，并不断提升数据质量。

(一)数据质量需求和管理目标

数据质量需求是度量和管理数据质量的依据，学校根据数据质量需求制定学校数据质量管理目标，以及管理维护围绕数据质量六个维度制定的校验规则和方法。

数据质量需求的形成。根据学校数据管理的目标、业务应用场景的需要、上级管理部门的要求，以及常见的数据问题来明确数据质量管理的范围，梳理各类数据的优先级，并参考相关的标准来形成数据质量需求，以及衡量数据质量的技术指标、业务指标。

数据质量管理目标的制定。学校需要根据自身业务管理的数据需求，建设数据质量管理目标，明确学校数据质量评估维度，制定数据质量评价体系和相关管理制度，明确数据人员权责和管理权限。

数据质量规则设计。依据数据质量需求和管理目标，定义各类数据的质量评价指标，校验规则和方法，并对这些规则进行规范化管理，以及持续的维护。

(二)数据质量检查

数据质量检查是根据数据质量规则中的有关技术指标和业务指标、校验规则与方法对学校的数据质量进行实时监控，发现数据质量问题并反馈给数据管理人员。学校应制定数据质量检查计划、建立数据质量问题管理机制，开展全面数据质量的监控。

数据质量检查计划。根据学校内外部的需要，制定数据质量检查的计划，明确数据质量检查的主要内容和方式，明确数据在各个阶段的检查点、检查模板，并建立数据质量检查制度、流程，强化数据质量检查的管理。

开展数据质量监控。根据业务的需要，建立相应工具，依据预先配置的规则、算法，对系统数据进行剖析和校验，帮助数据管理人员及时发现数据质量问题，掌握数据质量的真实情况。

数据质量问题管理。依托工具平台，建立数据质量问题的发现和告警机制，明确数据质量的负责人。对数据质量的问题进行管理包括问题的记录、查询、分发和跟踪。

(三)数据质量分析与提升

数据质量分析是对数据质量检查过程中发现的数据质量问题及相关的信息进行分析，找出影响数据质量的原因，形成数据质量报告，为后续提升数据质量提供参考。学校应对数据质量检查和问题处理过程开展有效分析，及时发现影响数据质量的潜在风险，形成并定期发布数据质量分析报告，提升数据质量，包括确定数据质量改进目标，制定并实施数据质量改进方案，持续优化和改进数据质量。此外，学校应建立数据质量问题知识库，完善数据质量问题分析方法，建立对数据质量的态度和共识，构建良好的数据质量文化。

二、数据安全

(一)数据安全策略制定

数据安全策略是数据安全的核心内容，学校应结合国家、教育系统等安全监管需求，认真分析梳理学校数据业务需求，全面分析数据安全风险，开展数据安全风险评估，制定相应的数据安全标准，规划建设数据安全策略。

数据安全策略和标准制定。职业院校应制定数据安全策略和标准，包括分析数据安全风险，开展数据安全风险评估，制定相应的数据安全标准和安全策略，明确数据安全管理的目标、原则、制度、组织、流程等，并根据内外部环境变化，定期修订优化数据安全标准和策略。

安全责任体系建立。教育行政部门和学校应对本地区、本部门工作中产生、汇总、加工的教育数据落实网络安全等级保护制度。按照“谁主管谁负责、谁运营谁负责、谁使用谁负责”的原则，数据所属业务的主管部门是数据的主管部门，应明确数据使用处理规则和防护要求；存储数据的信息系统主管部门是数据的运营部门，应制定信息系统安全准则，落实数据主管部门的防护要求；

利用数据开展业务的部门是数据的使用部门，应遵守数据使用处理规则和信息
系统安全准则要求。

(二)数据安全

在数据安全标准与策略的指导下，学校应清晰定义和细化数据安全需求，
基于数据安全标准和安全策略，明确数据安全等级划分，根据数据安全保护要
求落实数据安全风险管理，规定数据访问权限及权限控制，采用合适的信息安
全技术保障数据应用安全，进行用户身份认证和访问行为监控，重点数据安全
权限落实到字段级，明确核心字段的安全等级和管控措施，主动预防数据安全
风险发生，确保数据安全。

分级分类管理。学校应全面梳理本单位的数据，形成数据资源目录，准确
掌握数据基本情况，做到底数清、情况明。按照数据在教育发展中的重要程度，
以及一旦遭到篡改、破坏、泄露或者非法获取、非法利用造成的危害性程度确
定数据等级，对数据实行分类分级保护。数据分类可参考数据标准分类方式进
行。数据分级可以按照数据重要性维度，数据开放维度方面进行分级。数据重
要性维度可以分为一般、重要、核心三级。数据按照开放程度维度可以分为公
开数据、受限数据、敏感数据、涉密数据四个等级。

收集安全管理。学校采取合法、正当的方式依法依规采集数据，并保障数
据完整性、准确性和时效性，不得窃取或者以其他非法方式获取数据。新建系
统立项应对计划采集的数据进行分类分级，并对数据收集活动的合规性、合理
性进行审核；已建信息系统新增、调整数据收集活动也应履行审核程序。按照
“一数一源、最小必要”的原则，严格按照业务需要和职能边界确定数据收集
使用范围，优先通过共享获取数据，避免重复收集数据。面向师生收集信息应
公开收集使用规则，明示收集使用目的、方式、范围和存储期限。数据收集过
程中，应当根据数据级别采取相应的安全措施，加强重要数据采集人员、设备
的管理，并对数据的采集时间、类型、数量、频度、流向等进行记录。

存储传输安全管理。数据存储遵循“最短周期”原则，存储期限应当为实
现处理目的所必要的最短时间，超过期限的数据应进行归档或销毁。法律法规
对存储期限另有规定的，从其规定。在境内运营收集和产生的非公开数据原则
上应在境内存储，因业务需要确需向境外提供的，应按国家有关规定开展数据

出境安全评估。数据主管部门应按照数据类型、级别和数据特点，分别制定相应数据的备份策略，确保备份数据的正确性和完整性，定期开展数据恢复测试。需要长期保存的数据，应在介质有效期内进行转存，防止存储介质过期失效。数据主管、运营、使用部门应当根据传输的数据类型、级别和应用场景，制定安全策略并采取保护措施。传输重要数据时应当采取校验技术、密码技术、安全传输通道或者安全传输协议等措施。

开放共享安全管理。数据主管单位制定本单位的数据目录，明确共享属性和防护要求；信息化部门对共享需求和形式进行审核；数据使用单位遵守数据主管单位的防护要求，并接受数据主管单位和信息化部门的监督。数据公开与共享应完成数据公开共享审核以及对重要数据开放的风险评估，鼓励通过在线接口和“用而不存”的方式使用共享数据。

个人信息安全管理。个人信息是指以电子或者其他方式记录的能够单独或者与其他信息相互结合识别特定自然人身份或者反映特定自然人活动情况的各种信息，包括姓名、身份证件号码、通信通讯联系方式、住址、账号密码、财产状况、行踪轨迹等。学校应增强个人信息保护意识，充分尊重师生对个人信息的知情权和决定权。收集处理个人信息应以显著方式、清晰易懂的语言展示收集处理规则，并经个人信息主体同意后方可实施。个人信息的处理目的、处理方式和处理的个人信息种类发生变更的，应当重新取得个人同意。不得以默认、捆绑、停止安装使用等手段变相强迫授权，不得违反法律法规和超越约定来收集处理个人信息。利用个人信息进行自动化决策，应当保证决策的透明度和处理结果的公平合理。存储传输个人信息应采取加密措施，公开个人信息应采取去标识化处理。种族、民族、宗教信仰、个人生物特征、医疗健康、金融账号、个人行踪等信息属于敏感个人信息。收集处理敏感个人信息应对必要性、科学性、伦理性进行论证，并经学校决策同意后方可实施。实施时应取得个人信息主体的单独同意，并告知使用敏感个人信息的必要性以及对个人的影响。在公共场所采集图像、个人身份识别信息的，应为维护公共安全所必需，只能用于维护公共安全的目的，不得公开或向其他方提供。

使用处理安全管理。数据使用部门应明确数据操作人员的录入、查看、修改和删除等权限，利用数据开展统计分析、科学研究、决策分析时，应经数据

主管部门同意，并对数据进行必要的脱敏处理。处理数据的系统应具备身份认证、访问控制等技术措施，防止未经授权的数据活动，详细记录数据操作，相关日志保留时间不少于六个月。对用途不明确，存在安全风险隐患的数据申请，信息化部门可以不予提供。利用第三方平台或委托他人开展数据处理活动的，应由数据主管部门和被委托方签订数据安全与保密协议，明确使用方式、应用场景和使用边界，落实数据安全责任。委托处理重要数据的，应当对被委托方的数据安全保护能力、资质进行评估或核实，不得委托第三方个人进行数据使用与处理工作。

安全监测与应急管理。信息化部门建立数据安全风险监测机制，统筹建设数据安全监测预警技术手段，形成监测、溯源、预警、处置等能力，组织开展学校数据安全风险监测并与相关安全部门加强信息共享。开展有重大影响的个人信息处理活动应在事前提交学校进行风险评估；发生信息泄露、毁损、丢失等数据安全事件，或者发生数据安全事件风险明显加大时，数据主管部门应当立即采取补救措施，及时告知个人信息主体，并向信息化部门报告。

(三)数据安全审计

开展数据安全过程审计，确保数据安全目标、标准、策略、指导方针和预期结果相一致。开展合规审计，评估现有规范和规程是否适当，是否与业务要求和技术需求相一致，是否符合外部监管要求。开展供应商审计，评估确保供应商切实履行数据安全义务，及时发现数据安全隐患，及时改进数据安全管理工作；定期发布数据安全审计报告。

(四)数据安全技术支持

通过人员身份标识和鉴别、存取控制、审计日志、数据加密等技术的有效结合确保数据安全。对于敏感数据可采用数据脱敏技术实现在非可信环境下的使用问题。数据脱敏是指利用数据脱敏技术将数据库中的数据进行变换处理，在保持数据按需使用目标的同时，又能避免敏感数据外泄。

第六章 数据应用

一、数据分析

数据分析通过对内外部数据的深度分析与挖掘为学校各项业务管理活动提供决策支持，提高学校制定决策的精准度和科学性，提升学校的核心竞争力。

（一）对象分析

学生分析。围绕学生的基本情况、入学情况、学业情况、校园生活、校园活动参与情况、消费情况、图书借阅情况、身心素质、实习就业情况、竞赛获奖等信息形成学生画像和图表分析应用。

教师分析。围绕教师的基本情况、授课情况、教学情况、图书借阅情况、科研项目、科研成果、知识产权、社会服务、参与培训等信息形成教师画像和图表分析应用。

设备分析。围绕设备的采购情况、使用情况、维修情况、所在位置等信息形成设备画像和图表分析应用。

班级分析。围绕班级的学生整体情况、选课情况、成绩情况、授课教师情况、竞赛获奖情况、实习情况、就业情况、社团活动参与情况、消费水平、图书借阅情况、宿舍分布情况、网络使用情况、党团情况等信息形成班级画像和图表分析应用。

专业分析。围绕专业的学生整体情况、课程设置、师资队伍、教学改革、教学成果、科研水平、竞赛获奖、校企合作、实训基地、实训设备、学生学习与生活情况、学生实习就业情况等信息形成专业画像和图表分析应用。

院系分析。围绕学院或系部的学生情况、专业群设置、专业设置、课程设置、师资队伍、教学改革、教学成果、科研水平、竞赛获奖、校企合作、实训基地、实训设备、学生学习与生活情况、学生实习就业情况等信息形成院系画像和图表分析应用。

学校分析。围绕学校基本办学情况、师资队伍、人才培养、专业建设、教学科研、对外服务等信息形成学校画像和图表分析应用。

（二）主题分析

师资队伍主题分析。从师资结构、师资力量、教学科研、社会服务等方面进行分析支撑师资队伍的建设。

教学改革主题分析。从课程设置、课程资源建设、教材建设、教学评价、教学成果、教学竞赛、教学过程等方面进行分析支撑教学改革。

学生培养主题分析。对学生从德、智、体、美、劳五个维度进行数据分析，重点关注学生的学习动态和学业发展，及时掌握学生的培养进展。

科研情况主题分析。对科研单位和科研团队从科研队伍、科研项目、科研成果等方面进行科研发展情况分析。

财务决策主题分析。从财务收入、财务支出、固定资产、科研经费、预算收支、预算执行情况等方面进行分析支撑相关部门在财务运营、发展、偿还、预算管理方面的决策能力。

图书主题分析。对图书馆藏中纸质资源的借阅情况、数字资源的浏览情况进行分析形成热门图书和数字资源，对阅览室座位使用率、人员出入情况等分析优化资源配置。

消费主题分析。对学生消费情况、各区域消费情况、各时段消费情况进行分析支撑受欢迎餐饮推荐、贫困生推荐、错峰就餐管理等应用。

党员学习情况主题分析。对党员基本信息、党员干部学习、三会一课、党员日常、党员主题党日等主要数据指标，展现党员的学习表现、思想水平、党性修养等情况。监督反馈党建活动的开展频次和质量，提升院校基层党建质量。

（三）综合分析

常规报表综合分析。学校日常管理中存在大量的数据报表，如高基报表、人才培养状态报表、“双高”建设报表等，需要对数据进行综合分析并按照规定的格式输出和展示，以满足上级教育管理部门及学校各部门对各类报表的需求。

多维综合分析。对教师、学生、教学、实习实训等相关的业务关联性较大的数据进行综合分析，通过对不同对象、不同维度数据的对比分析、关联分析等梳理和归纳各类数据之间的度量关系，帮助教学与管理人员及时发现业务管理中的盲点和漏洞、探索与分析影响管理效率和效果的关键影响因素等。

二、数据服务

数据服务是通过对学校内外部数据的统一加工和分析，结合教学与管理的需要，以数据分析结果的形式提供跨部门、跨学校的数据服务。学校应统筹考虑构建基于数据的应用服务平台以及相应的工作机制，探讨数据服务模式和内容，充分发挥数据价值，在保障数据服务的质量、有效和安全基础上，满足不同类型用户和应用的数据需求。

学校应充分利用大数据中心的数据资源，为师生员工提供体验更好、功能更丰富的服务，主要包括四个方面的服务：党建引领、人才培养、学校治理、社会服务等。

（一）党建引领

职业教育作为培养高素质技术技能人才的重要教育类型，肩负着党建育人的重担。教育大数据能够在个性化党性教育、基层党建质量提升、党员考核方式创新发展三大场景发挥作用。

个性化党性教育。支持各级党组织及时发现党员学习问题，针对性地提供相关学习资源和学习指导；支持党员查看学习进展，掌握学习进度，根据学习情况及时调整自身学习状态。

基层党建质量提升。掌握基层党组织活动开展和所属党员思想动态的情况，对活动开展不足的基层党组织实施预警，促进基层党组织党建活动的开展频次和质量，升级和优化队伍建设。

党员考核方式创新发展。通过行为数据采集和画像分析，实现对党员的智能监督考察和预警。具体体现在以下三个方面：

- 党员画像，支持党员自查、自省、自我批评，并公开优秀党员画像发挥“榜样效应”，支持新党员向优秀榜样学习。
- 记录入党积极分子发展轨迹，形成发展图谱，支持党组织有针对性地发展和培养新兴力量。
- 综合民主评议、党员自评、党员互评、入党联络人反馈、思想汇报、党建活动参与等数据，构建一体化的考核评价体系。

（二）人才培养

人才培养是职业教育的最核心场景。教育大数据能够应用在师生发展、数字资源、信息化教学、实习实训、专业建设等场景，提升职业教育人才培养质量。

1. 师生发展

学生画像。多维度呈现和勾勒学生成长图谱，助力学生个性发展。具体体现在以下三个方面：

- **个性化教育方案。**记录学生从入校到毕业就业的全过程成长数据，全方位、立体化地了解学生，并结合学生特点制定和调整个性化教育方案。
- **综合素质评价。**采集学生发展状态数据和行为数据，了解和监测学生发展状态，及时掌握学生变化趋势。依据学生综合素质评价指标体系对学生进行量化考核和综合素质评价。
- **学情分析与预警。**采集学生学习行为数据和结果数据，生成学情分析报告，揭示其出勤情况、在线课程登录情况、课堂活动参与情况、协作学习互动情况、课程资源查阅情况、课程任务完成情况等，进而指导学生进行自我反思、动态调整学习目标和进度等。对低于教学预期的学生，发出学习状态、学分缺失预警标识来提醒学习者和教学者，以此增加其课程保持率、提高课程通过率。

教师画像。多维度地呈现和勾勒教师发展图谱，助力教师良性发展。具体体现在以下两个方面：

- **全面记录教师从招聘、报到、注册、工作、离校、退休的全过程在职服务数据，**刻画教师职业全生命周期，帮助教师制定个人职业生涯规划 and 职业成长方案。
- **采集教师发展状态数据，支持从多维度、多元化对教师进行评价和考核，**构建全方位、过程性、人性化的评价体系，助力教师全面发展。

2. 数字资源

数字资源推荐。围绕课程相关的信息推荐各平台的精品课程资源、图书馆的数字资源等。根据学生所学的课程与成绩、参与的活动、个人基本信息推荐社团活动、学术讲座、选修课程、竞赛活动等。

数字资源建设。通过分析各专业、各类别资源建设数量和质量的情况揭示学校数字化教育资源建设现状，根据学校数字化资源的应用情况揭露资源的供需现状，从而预测未来数字化教育资源的应用趋势和走向，助力学校进行针对性地资源补给，优化数字资源的供给体系。

3. 信息化教学

混合课程设计。整合课程建设数据和教学过程行为数据，帮助教师了解自身教学表现与混合课程建设情况，发现教学问题，反思并改进教学过程，优化不同教学情境中的教学设计。此外辅助教学管理者洞悉各院系、各专业的混合课程运行情况，有助于调控教与学进度。

教学干预。通过对在线行为数据和课堂行为数据的采集与分析，实现在线自动干预和人工干预相结合，提高干预精准度并完善效果。

教学评价。基于学生学习行为数据和课程教学数据扩展教学评价方式，支持将评价融入到教学全过程中，并将评价阶段相对前移；关注学生的学习态度和行为投入情况，改变以往只通过试卷、作业成绩和课堂表现等方面来评价学习效果的情形；加强课堂互动，实现多种形式的即时测评和同步反馈，为及时做出教学干预和评价等提供依据，也为学生的自我评测和自主调节提供帮助。

4. 实习实训

实习实训过程支持。通过对学生实习实训全过程的跟踪和动态监管，支持学校和企业导师及时掌握学生实习实训进度和在实习实训中遇到的问题和困难，和学生有效沟通和交流，并提供必要的指导和帮助。

实习实训评价方式。基于实习实训过程数据，促进实习实训评价方式转型。一方面，可结合企业导师对学生在岗位实习实训中的定量评价数据和学校导师对学生的定性评价数据，精准定位学生的实习实训情况；另一方面，通过对学生实习实训全过程数据的记录，支持兼顾对实习实训过程和结果的评价，打破以往只注重结果的评价方式，关注学生在实习实训过程中的行为表现。

5. 专业建设

院校专业结构调整。综合区域经济和产业发展相关数据，预测区域产业发展趋势和岗位变化情况，并据此调整院校专业设置。具体可通过设置新兴专业和升级传统专业，适应大数据、云计算、人工智能等新兴信息化技术对传统产业产生的变革影响，加强专业设置与社会需求的契合度。

专业人才培养方案调整。综合市场需求和毕业生去向等数据，揭示企业对技术技能人才的需求，整合学校专业优势，利用数据分析专业课程建设与职业岗位要求匹配度，从而动态调整专业人才培养方案和专业课程设置，提升教学内容与岗位实际需求的契合度。

（三）学校治理

学校治理包括管理、服务、决策等。教育大数据可在信息化管理、信息化服务和科学化决策等场景发挥重要作用，提升院校治理能力和师生办事体验。

1. 信息化管理

校园安全监控和预警。整合 GIS 的综合监测平台、网络数字视频监控系统、智能交通管理系统、消防一体化系统、校园网格化管理系统、周界防范预警系统、日常巡视检查系统、门禁系统等相关信息，形成“精准督察、过程留痕、决策有据、指挥联动、预警为主、主动干预”的校园安全垂直管理体系，实现校园安全的统一调度、多级监管、联动处理。通过大数据分析和人工智能技术对安全风险进行预测和预警，实现安全管理工作可量化、可视化和可控化，为校园安全业务开展提供科学的辅助决策，使校园更安全、更舒适、更高效。

学生失联告警。通过对学生的学习、消费、上网、上课、刷卡、图书馆、宿舍等各方面的数据进行建模分析，判断学生在校状态，检测异常情况，并提醒辅导员和班主任对异常状态的学生重点关注，并及时干预。

2. 信息化服务

资源优化配置。基于资产设备、教学环境、场馆的建设和使用情况，掌握教学、校园生活等对资源的需求状况，以及资源供给与需求的匹配情况，为科学制订资产建设规划、优化资源配置、提高资源利用效率提供决策支持。

就业指导。预测学生毕业去向，为学生提供就业的精准指导。具体体现在以下三方面：

- 全面收集毕业生在校日常学习、竞赛、实习实践等信息，通过对数据的收集、分析和挖掘，对比往届毕业生的毕业去向，构建毕业生完整活动数据链，精准描绘基于人格、能力、职业规划等维度的“学生职业画像”，实时预测在校生未来毕业去向，为学生提供就业决策信息。
- 聚焦企业用工和岗位需求信息，定向采集优质企业的就业资源，强化招聘资源整合，分省份、区域、行业、层级定期向学生精准推送各专业岗位的需求情况，便于打通相关就业主体之间的信息传输。此外，充分整合学生职业能力、职业画像、精准就业等系统信息，为毕业生和岗位需求精准“画像”，实现供需全面对接，减少大量中间环节，为企业和学生提供便利。
- 通过数据分析，对毕业生电子签到、简历完成度、线上简历投递、录用通知数据比对等就业行为数据深度挖掘，精准分析学生求职意愿和行为数据，对学生针对性地进行就业辅导。针对其中存在就业困难的学生群体，提供就业指导服务，开展职业生涯规划教育，精准推送人职匹配度高的工作岗位，实现“一生一策”，攻克求职薄弱点，提升就业成功率和满意度。

资助帮扶。基于一卡通消费、学生家庭信息、生活学习情况、历史资助信息、资助获得后消费习惯改变等数据精准识别贫困生、量身定制“菜单式”资助方案，从而实现对贫困生的精准资助、精准育人和精准评估。

3. 科学化决策

校情分析。全面汇聚教学、科研、招生就业、后勤保障等各个业务系统的多维度数据，通过综合仪表盘的方式对学校各方面数据进行统计、分析、预警和预测，提供具有扩展性的、准确的、多维度的学校建设和发展情况可视化服务，结合管理者的查询情况对数据进行个性化的展示。一方面支持学校管理者从海量信息中快速获取准确、直观的信息，为领导层决策提供支持；另一方面通过综合分析对学校各类业务数据指标的计算与分析，衡量学校人才培养质量等工作。

学校战略规划。基于校情分析诊断当前在办学水平、办学效果等方面存在的差距和不足，因校制宜制定科学合理的战略规划，促进学校全方面事业发展。

绩效评价。建立大数据分析 KPI 系统，评价教学质量、专业建设水平、教师绩效、学生综合素质等事项，建立学校质量效益评价体系，从投入、产出、过程、结果等四个方面科学准确地检测、评价、指导、考评学校的办学效益，为学校高质量发展提供保障。学校管理者依据绩效评价结果科学地开展教学管理与督导，实现学校、专业、课程、教师、学生五个层面的预警分析和智能决策，通过优化流程来实现增值并促进高效管理。

（四）社会服务

职业教育承担着社会服务的重任，教育大数据能够在技术服务、校企合作、校友联络三大场景发挥作用。

1. 技术服务

基于科研相关数据，将外部技术需求数据库和院校内部科研成果库进行对接和匹配，加快科研成果被认知及推广和应用，发挥科研成果的社会和经济效益，助推产业升级。

2. 校企合作

基于校企合作信息、就业信息，整合分散于政府各部门、各行业组织、企业，以及职业院校的各类信息资源，支持多方融通的校企合作立体化信息发布，支持面向学校、面向企业、面向行业的多模式应用支撑服务，支持形成“互联网+校企合作”新模式，促进校企深度融合。

3. 校友联络

为每个毕业学生建立校友档案，划分校友群体，提升校友资源圈质量，加强校友联动。辅助校友的招聘工作，直观反映校友企业招聘状况，帮助校友企业全面了解准毕业生，为岗位设置和聘任提供科学决策。

第七章 大数据平台建设

一、支撑环境建设

大数据中心支撑环境描述大数据中心建设所应该具备的硬件和网络基础设施等内容。

（一）存储计算资源

建立融合计算资源池，如虚拟机资源池、容器资源池、物理机资源池，并充分利用可用计算资源，纳入利旧资源池统一管理。数据计算资源用于承载全校各类信息系统的服务器需求服务，提供需要的计算资源及存储资源：虚拟机计算池主要为一般性应用提供虚拟服务器需求；容器资源池主要为微服务等新应用提供容器需求；物理机资源池主要为特定需要较高性能的应用提供物理服务器需求；利旧资源池为测试、实验类业务提供相应计算资源。

建立数据业务持续性计划管理体系，并建设实施相应的容灾备份、业务接管等保障体系。职业院校可根据自身实际情况实现灵活的备份方案，例如：建立本地灾备方案，通过备份容灾软件实现数据业务层面容灾备份；建立基于存储快照的灾备方案，通过存储之间的块复制快照方式实现底层数据层面容灾备份；建立基于双活存储的同城灾备方案，基于数据内部存储设备之间的双活机制实现实时的同城业务双活；建立基于 CDP 的异地灾备方案，部署 CDP 等设备，基于 CDP 的持续数据保护机制实现异地灾备。

（二）网络安全环境

大数据中心始终牢记网络安全职责与使命，认真履职。从物理访问控制、基础设施安全、网络架构、边界防护等方面加强，建立了安全管理中心，对设备状态、恶意代码、补丁升级、安全审计等安全相关事项进行集中管理。以网络监测为基础，实时发现网络边界和关键安全区域的威胁；提供特征检测及异常检测等多种监测技术手段全覆盖监测各类安全事件；提供对网络攻击事件的监测能力，识别网络攻击与异常信息。重点做好访问控制，实现基于 IP 地址、MAC 地址、主机名、操作系统用户名、应用程序名称和账号信息等多个维度对用户身份进行鉴别。

二、平台技术框架

大数据中心平台技术框架描述了服务大数据中心建设所需的技术结构和工具（如图 3 所示），采用分层结构，包括数据源层、数据汇聚层、数据存储与管理层、数据分析与开发层、数据应用层、以及数据治理监控层。

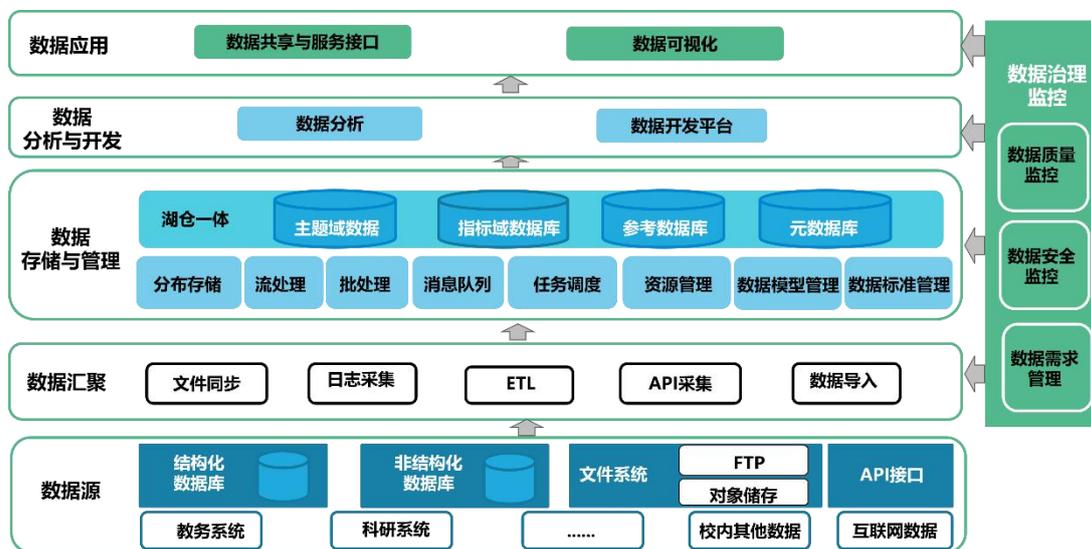


图 3 大数据中心平台工具框架图

（一）数据源层

数据源层：数据源包含校内外两方面数据：校外数据主要指互联网数据，如微信、微博、贴吧、各教育网、政府网等各领域的的数据；校内数据主要是各部门应用系统所产生的业务数据（如：教务系统、办公系统、财务系统等）及校内其他数据（如：上网行为数据、监控视频流数据等）。各个数据源存储在主流关系型数据库、数据仓库、大数据存储系统中，形成教育大数据分析的基础资源。

（二）数据汇聚层

数据汇聚作为底层基础支撑性服务，是大数据环境的基础组成部分。通过数据集成工具对学校的多源异构数据汇聚需考虑不同类型（结构化数据、半结构化数据、非结构化数据）、不同集成策略（定期采集、实时采集）的数据汇聚方案，并支持数据的预处理，为大数据环境提供原始数据支撑。需提供异构数

据汇聚工具，操作者借助该工具可以通过流程图式的图形化工具快速、灵活地设计 ETL 过程，并能方便地进行部署、调度及监控数据等管理活动。

利用 Datax、Kettle 等主流的 ETL 工具将来源于不同存储系统的结构化、非结构化和半结构化的数据汇聚到大数据平台。需支持 MySQL、Oracle 等对事务性要求较高的关系型数据库，支持 HDFS、Hive、HBase、Greenplum 等大数据存储系统，以及支持 MangoDB、Redis 等 NoSQL 数据库。数据汇聚过程须任务化，支持任务异步在线运行、定时运行，能查看任务运行日志和结果。

（三）数据存储与管理层

数据存储与管理层是大数据处理环境的核心，它存储由数据采集层采集回来的各类数据和数据治理后的各个层次的数据，并为上层应用提供数据处理的能力。

数据存储主要完成多源数据及多类型数据（结构化、半结构化、非结构化）的融合存储，并基于分布式关系型数据库系统、分布式文件系统、分布式内存数据库来实现不同层次数据的存储。包括通过分布式关系数据库实现数据仓库各分层数据、元数据、参考数据、系统管理与配置等数据的存储管理；通过分布式文件系统实现数量较大的过程数据、半结构化、非结构化数据的存储；通过分布式内存数据库实现服务与应用数据的缓存。数据处理支持数据的批处理和实时处理。批处理是一种离线数据处理模式，应支持分散-聚集的处理方式。实时处理包括流处理和内存计算，流处理支持滑动窗口式的实时分析任务，并提供容错机制以实现故障恢复。内存计算通过 DAG 执行引擎提供不同类型数据的处理能力。数据管理是利用软件系统工具支持数据元管理、参考数据管理、指标体系管理；支持数据标准的生命周期管理；支持根据数据标准进行主题域模型建模与设计。模型建模与设计主要依据主题域、数据元、维度、指标、度量构建数据仓库各类不同层的数据表；通过在数据表依据绑定的数据元来进行表的建设确保了数据标准的统一，并为数据质量管理提供支撑；通过定义数据

表中各数据项与维度、指标的关联，实现数据表数据间的关联，为服务应用层提供数据间的穿透与关联查找。

数据存储与管理层可利用 Iceberg、Hive 等主流技术构建湖仓一体的大数据存储平台，形成分布式的、高容错、高稳定的结构化、半结构化、非结构化的数据存储体系，基于 Hadoop、Spark、Flink 等主流运算架构，以 Kafka，Azkaban，Yarn 等主流技术框架作为队列、任务管理、资源协调的工具，形成稳定快速的、可规模化的统一计算平台，支持对海量多源异构数据的存储、处理与快速查询。数据存储与处理层应采用能支持对分布式主机集群的各种计算服务资源统一管理调度，主要的功能应包括：能实时监控集群中各个主机内存、磁盘、CPU 等状态；能在线快速无缝扩展、替换不同功能的主机；能灵活在集群中增加常见的大数据服务，如 HDFS、Hbase、Hive、Flume、Hue、Kafka、Spark、Sqoop、Zookeeper、Yarn、Oozie、Impala 等。

（四）数据分析与开发层

数据分析与开发层提供对数据的探查与自主分析和图形化的数据开发。

数据分析工具。提供对常用关系型数据库（如 MySQL、Oracle、SQLServer）和常见数据仓库（如 Hive）的数据字段的探查。提供不同类型数据库中数据的快速查看、表结构提取、基于 SQL 数据查询等自主分析功能。维护查询分析的脚本、运行状态等历史记录。

数据开发平台。是可视化的低代码工具平台，支持通过可拖拽的方式实现丰富的数据同步、数据处理、数据分析与挖掘等功能，帮助学校降低数据开发门槛，提高开发和运维效率，提升数据资产价值。数据开发平台支持大数据存储运算、人工智能模型调用等各种任务的编排、调度和监控；支持 Shell、Perl、Python 等脚本资源的管理和在线运行与调度；具备任务管理功能，对任务流程修改和版本变更进行管理；具备详细的异常日志记录，能对错误进行定

位，通过监控调度运行中的所有运行日志信息，提升系统开发和维护的质量，减少系统建设、实施及运维等各方面成本。

（五）数据应用层

基于数据仓库理念封装不同类别的数据，以 Restful API 的方式对外为不同应用提供统一的数据共享服务接口。针对具体的分析业务场景，组织各类数据资源和服务工具，建立基于大数据技术的业务应用，解决具体的业务问题，不同的应用具有各自的分析功能和展示界面，例如：数据穿透查询、教学科研主题分析、学生上网行为分析、学校舆情分析等。

数据共享与服务接口。以接口访问的形式提供数据资源服务，实现应用和数据分离。数据资源应封装成 Restful 等主流架构对外提供服务，基于标准化接口，实现对不同范围、不同授权的数据访问的 API 封装，以便于上层应用进行增值开发利用，体现数据价值。数据接口应提供实时统一的数据访问入口，屏蔽异构数据的复杂性，同时控制数据访问与数据内容的安全性。具体功能需包含服务概览、服务管理、策略管理、以及服务管理配置及相应的事件跟踪记录。服务管理支持 API 服务的创建、授权，调用的情况统计，并提供相应的应用和 API 集合的管理功能；策略管理支持对 API 服务进行流控制、访问控制的管理，并提供报警功能。

数据可视化。支持将数据结果以图文方式构建成不同分析页面的灵活平台，具备图形组件、页面编辑、动态数据源等功能。图形组件能将数据服务接口提供的数据进行可视化的展示，拥有一定的数据筛选与分类功能。提供丰富的图形组件库和组件自定义功能，包括面积图、关系图、雷达图、条形图、趋势图、饼图、折线图、柱状图、表格、散点图、漏斗图、水球图等。页面编辑能支持通过图形化拖拽形式将不同的组件进行组装构建成满足个体分析、主题分析、综合分析的页面，并实现数据在不同页面中的穿透；页面能支持按照不同类别、不同时间、不同人员范围对数据进行分析。动态数据源支持通过配置和脚本代码实现可视化图表的数据绑定，数据类型支持关系型数据库、数据仓库、静态数据、Excel、API 接口等多种数据源，并能对数据源进行管理实现多种数据来

源新增，编辑，删除，查看管理功能。通过数据可视化可构建报表的 UI 界面和展示大屏，并支持拖拽组件栏任意组件，快速构建不同类型的数字驾驶舱。通过交互式实时数据可视化应用构建针对不同实体、不同业务主题的画像，以生动、炫酷的形式即时呈现隐藏在庞杂数据背后的价值。

（六）数据治理监控层

数据质量监控平台。需围绕数据的完整性、一致性、准确性、有效性和及时性，进行数据监控，提升数据质量，并从数据接入、数据加工等环节，评估数据与需求的结合度与时效性。质量监控须提供数据质量大屏和数据质量报告功能、支持通过事前定义数据的监控规则、事中监控和控制数据的生成过程、事后数据评估和问题追溯过程，保障高质量的数据。监控的规则应支持常见空值、邮件、IP 地址、经纬度、唯一值等通用规则模板的定义，支持按照表或字段的规则类型分类、数据校验类型、数据表来源数据库的类型进行分类管理。监控规则的管理除增删改查外，提供试跑功能。监控执行需具备完善的历史记录，能对执行的监控实例进行溯源、查看执行结果和相关日志；根据时间范围、监控结果进行分类，监控执行情况。数据质量的监控最终以数据质量报告的形式呈现，报告应详细阐述近期综合质量得分的情况及趋势。

数据安全监控平台。需围绕着数据生存周期内的安全问题，提供全方位的安全管控能力，包括支持常用的脱敏算法、具备脱敏规则的配置、实现资源访问的授权管理、用户管理、角色管理、数据授权管理、报警功能的设置。权限管理支持数据授权须基于角色和数据资源的授权，且粒度能细致到数据服务的每一个接口，并根据角色或岗位与不同数据资源的关联情况形成用户与资源接口的数据授权管理。

数据需求管理。数据需求管理是对不同用户在获取数据服务的过程中所提出的数据需求进行管理，具体内容包括了数据需求提出、方案审议、数据开发、数据授权、结果回复等流程节点的管理。

三、建设模式

大数据中心的数据规划、建设、管理、应用和服务等内容应体现学校主体地位，相关内容、架构、技术路线基本一致。因此，大数据中心建设模式主要从环境建设和运行管理的角度，提出三种建设模式：自主建设运营管理的模式、区域建设模式以及校企合作建设模式。

（一）院校自主建设模式

院校自主建设模式指由学校自主建设和运维大数据中心环境，包括核心机房、网络基础设施、计算资源、数据存储资源、结构化数据库系统、非结构化数据系统以及数据集成与共享平台系统，各类数据应用与服务软件平台，以及信息网络安全设施等。此外，院校还需要建立独立的异地（至少是不同楼宇）数据容灾环境，包括机房环境、计算资源、数据存储资源以及配套的系统平台。

院校需要建立配套的运维体系以保证自建大数据环境的运行，包括专业的运维队伍与人员、标准化的运维管理制度与工作流程。

此外，院校需要建立一定规模的专业化的数据管理及应用开发技术队伍，负责支撑学校数据建设与管理的，包括数据规划、数据架构、数据治理、数据集成共享、数据质量、数据全生命周期、数据标准、数据安全等方面的技术性工作。

（二）政府主导的区域建设模式

政府主导建设模式指由政府主导建设和运维集中式的大数据中心环境，以云计算、云服务模式向区域内院校提供环境和服务。

政府负责集中建立大数据中心环境、运维体系、安全体系以及异地容灾备份环境。不同院校共同使用大数据中心环境和服务，但各自的多租户空间、互不干扰。

此种模式下，可最大程度地优化配置资源，减少院校的投入，同时有利于标准规范的形成，有利于建设长期化可持续发展的大数据环境，有利于加强院校间的合作与交流，特别是对于资源匮乏和技术能力欠缺的院校，能够大幅度降低院校建设和使用大数据中心的门槛。

需要强调的是，即便是政府主导建设大数据中心，院校仍然是大数据建设管理的主体，仍然需要数据建设和管理纳入自身的规划和战略行动，仍需建立

数据建设、管理与应用相关的队伍、制度和工作体系，仍需根据院校自身的需要开展数据规划、数据治理等各方面的工作。

（三）校企合作建设模式

校企合作的建设模式，由企业提供大数据环境，由院校独立或者和企业合作运营数据中心。

校企合作的大数据中心模式可以促进院校以大数据课程体系建设、大数据实验室建设、大数据师资建设为切入点开展产教融合、校企合作探索应用型大数据人才培养模式。

第八章 全国职业教育智慧大脑院校中台对接

全国职业教育智慧大脑院校中台是国家职业教育智慧治理体系的重要组成部分，通过国家、省级、市、县区、学校信息管理平台汇聚实时数据，促进数据的跨层级、跨地域、跨部门有序流动。运用云计算、大数据等新技术加大数据分析应用的力度、深度和效果，通过可视化界面进行智能管理，实现个性化、精准资源信息的智能推荐和服务，为管理人员和决策者提供及时、全面、精准的数据支持，逐步形成“数治教育”新体系，提高决策的科学性。

一、数据对接内容

“全国职业教育智慧大脑院校中台”将通过数据中台联通教育行政部门和职业院校，覆盖部、省、市、县、校五级，实现学校、专业、教师和学生等全方位的数据直连、信息直达，建设“中国职教大脑”，服务教育治理。“全国职业教育智慧大脑院校数据中台”对职业院校的学校概况、教学管理、教职工管理、学生管理、党建思政、资产与设备管理、科研管理、服务管理等八个方面提出了数据标准化的要求。高职院校与中职学校的数据标准要求参见《全国职业教育智慧大脑院校中台高职数据标准及接口规范》和《全国职业教育智慧大脑院校中台中职数据标准及接口规范》。

二、数据对接方式

全国职业教育智慧大脑院校中台数据对接，采用标准化数据接口推送模式。接口设计遵循RESTful模式，各院校可通过数据接口采集API，按照数据对接接口开发规范，实现学校数据上报。

三、数据对接技术规范

（一）数据对接流程

数据对接流程如图4所示：

1. 院校可通过数据上报资质申请页面申请对接账号信息，如之前已申请过账号，只需通过资质申请页面补全信息即可，申请成功之后相关账号信息及接口加密方式通过邮箱方式进行下发，用于接口调用上报数据。
2. 院校调用用户授权接口，获取接口调用需要的令牌信息。用于后续的数据上报和查询操作。

3. 院校调用数据增量采集接口，将相关业务数据上传到全国职业教育智慧大脑院校中台中，系统会对上传的数据质量进行检测，将检测结果反馈至试点院校，院校可通过授权的账号信息登录全国职业教育智慧大脑院校中台进行查看。

4. 院校可调用增量数据上报查询接口来检验上传数据是否符合规范，不符合规范的数据会被清洗，需要检查调整后重新上传。

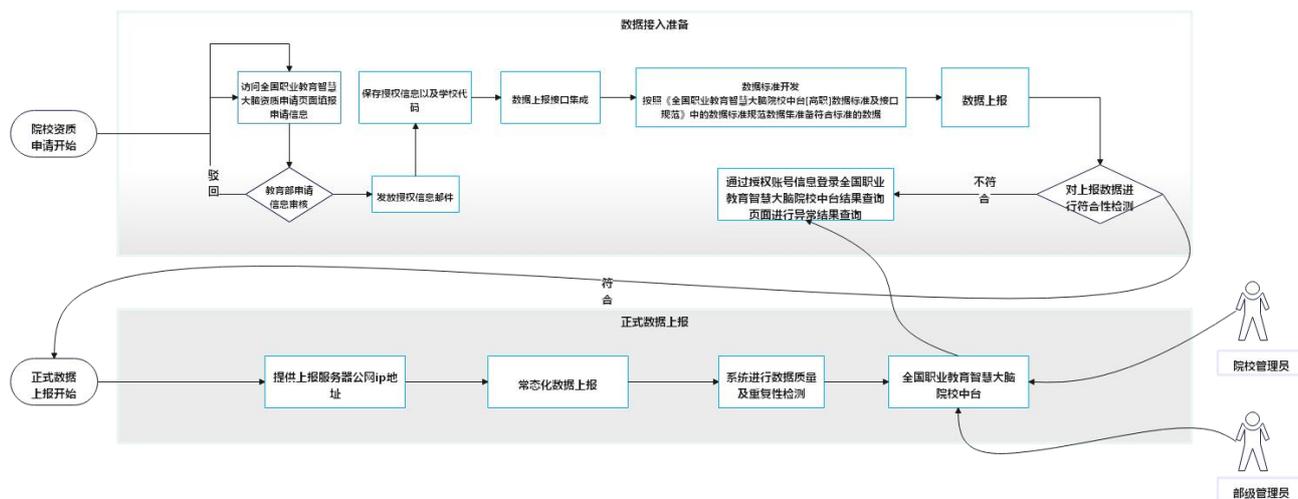


图4 数据对接流程图

（二）数据接口规范

数据对接接口调用规范包含用户授权接口、用户授权令牌刷新接口、数据上报接口、数据上报结果查询接口等4个数据上报接口，明确了接口请求方式、输入参数描述、输入参数示例、输出参数描述、输出参数示例等内容。

详细数据接口规范请参考本章第五节数据对接接口开发规范。

（三）数据安全要求

1. 院校在数据上报前需要提供调用数据上报接口外网IP地址进行白名单配置，配置白名单的IP才可以调用上报接口。

2. 院校在进行数据上报时需对学生身份证号、教师身份证号等敏感信息进行加密后再上报。

四、数据对接准备

（一）对接账号申请

院校可通过数据上报资质申请页面，填写申请学院名称、申请人、申请说

明、联系电话、联系邮箱、学校统一社会信用代码、学校类型等信息，进行账户资质申请，申请后等待管理员进行审核，审核通过后会以邮件的形式发送数据上报相关账号信息。

（二）数据整理

根据《全国职业教育智慧大脑院校中台高职数据标准及接口规范》和《全国职业教育智慧大脑院校中台中职数据标准及接口规范》结合院校实际情况进行数据整理，包括每个数据子类对应的学校业务系统、数据上报频率要求。

（三）数据标准开发

按照《全国职业教育智慧大脑院校中台高职数据标准及接口规范》和《全国职业教育智慧大脑院校中台中职数据标准及接口规范》要求开发准备符合标准的数据。数据上传后，系统会对数据格式、基础数据标准一致性进行检测，结果一般会在30分钟内输出，院校可通过授权的账号信息登录全国职业教育智慧大脑院校中台进行查看。

（四）数据接口集成

接口集成开发请参见第三节接口集成开发文档。建议院校通过已建设的校本数据中心进行数据标准开发和接口集成，实现接口自动上报。

五、数据对接接口开发规范

（一）接口调用方式

1. 接口授权

每次上报数据前，需调用用户授权接口，获取app_key和app_secret，用作接口鉴权。一次完整数据上报，获取一次即可。

2. 接口调用-请求体（BODY）

参数值：将接口参数值封装为JSON字符串，使用app_secret对JSON字符串进行加密。对密文重新封装为 JSONObject，以cipherText 为键，以密文为值。

加密方式：SM4加密算法。

3. 接口调用-请求头（Header）

app_key	应用授权校验码，通过调用用户授权接口获取。
timestamps	接口调用时客户端系统时间戳，单位：毫秒。参数值与国家授时中心标准时间戳比对差值不能大于 30s。
random	随机值，是客户端随机生成的值，作为参数传递过来，随机值的目的是增加 sign 签名的多变性。随机值一般是数字和字母的组合，6 位

	长度，随机值的组成和长度没有固定规则。
sign	接口签名，通过 BODY 中加密前参数 JSON 字符串+timestamps+random 进行 SM3 加密生成。

请求头键值说明：调用接口时，需要将app_key参数写入请求头（Header）中，后端服务获取请求头中的app_key参数值，进行授权校验。

校验不通过的情况：请求头中app_key参数不存在；app_key参数值处于未授权状态。

4. 接口请求数据签名规则

调用接口前，需要使用Body中加密前的上报数据JSON字符串，以及timestamps、random 参数值拼接成字符串进行 SM3（摘要算法）加密，形成sign参数值。

验签失败可能的情况：

- (1) 请求头中sign参数不存在。
- (2) 端对请求参数加密值与请求头签名值（sign）比对不一致。
- (3) 请求头中timestamps参数不存在。
- (4) 请求头中获取timestamps参数值与国家授时中心标准时间戳比对差值大于30s。

（二）数据对接接口调用规范

数据对接接口调用规范包含用户授权接口、用户授权令牌刷新接口、数据上报接口、数据上报结果查询接口等4个数据上报接口，明确了接口请求方式、输入参数描述、输入参数示例、输出参数描述、输出参数示例等内容。

1. 用户授权接口

接口描述：用户授权接口

请求URL：

http://ip:port/prod-api4/api/web/auth/login/interface/oauth

请求方式：POST（参数在表单(x-www-form-urlencoded)中）

请求参数：

参数名	必选	类型	说明
grant_type	是	string	授权类型：password
client_id	是	string	客户标识
username	是	string	用户名

password	是	string	密码
----------	---	--------	----

关键返回参数说明：

参数名	说明
access_token	令牌
refresh_token	刷新令牌
expires_in	令牌有效期
refresh_expires_in	刷新令牌有效期
client_id	客户标识
scope	身份
openid	openid
client_secret	刷新令牌密钥
app_key	应用 key
app_secret	加密密钥

返回示例：

```
{
  "code": 200,
  "msg": "ok",
  "data": {
    "access_token": "xxxxx",
    "refresh_token": "xxxxx",
    "expires_in": 7199,
    "refresh_expires_in": 2591999,
    "client_id": "xxxx",
    "scope": "",
    "openid": "xxx",
    "client_secret": "xxxx",
    "app_key": "xxxx",
    "app_secret": "xxxx"
  }
}
```

2. 用户授权 token 刷新接口

接口描述：用户授权token刷新接口

请求URL：

```
http://ip:port/prod-api4/api/web/auth/login/interface/refresh
```

请求方式：POST（参数在表单(x-www-form-urlencoded)中）

请求参数：

参数名	必选	类型	说明
grant_type	是	string	处理类型：refresh_token
client_id	是	string	客户标识：1
client_secret	是	string	刷新密钥
refresh_token	是	string	刷新令牌

关键返回参数说明：

参数名	说明
access_token	令牌
refresh_token	刷新令牌
expires_in	令牌有效期
refresh_expires_in	刷新令牌有效期
client_id	客户标识
scope	身份
openid	openid

返回示例：

```
{
  "code": 200,
  "msg": "ok",
  "data": {
    "access_token": "xxxx",
    "refresh_token": "xx",
    "expires_in": xx,
    "refresh_expires_in": xx,
    "client_id": "xx",
    "scope": "",
    "openid": "xxx"
  }
}
```

3. 数据上报接口

接口描述：按数据上报时间段查询数据上报结果

请求URL：

```
http://ip:port/prod-
```

api4/api/web/convergence/collect/interface/saveInterfaceData

请求方式：POST（参数在BODY，JSON格式）

请求头：

参数名	必选	类型	说明
Authorization	是	string	令牌
Content-Type	是	application/json	请求类型
app_key	是	string	应用授权校验码，通过调用用户授权接口获取
timestamps	是	string	接口调用时客户端系统时间戳，单位：毫秒。参数值与国家授时中心标准时间戳比对差值不能大于 30s
random	是	string	接口调用时系统随机码，可自行生成
sign	是	string	接口签名，通过 BODY 中加密前上报数据 JSON 字符串+timestamps+random 进行 SM3 加密生成

请求体加密参数：

参数名	必选	类型	说明
dataObjName	是	string	数据对象名称，详情见附件全国职业教育智慧大脑院校中台高职与中职数据标准数据子类
fileds	是	List	数据项集合

请求体加密后参数：

参数名	必选	类型	说明
cipherText	是	string	值为密文

请求体加密参数示例：

```
{
  "GZZYQKSJID": "ahjxxxjd_xxx",
  "XXJGDMn": "3634000467",
  "XXJGMCn": "xx 学校",
  "DZZMC": "第八党支部",
  "DZZBHn": "8",
  "HDXSn": "研讨",
  "DKZJRn": "王涵",
  "HDNRn": "xxxx",
  "HDKSSJn": "2022-09-28",
  "HDJSSJn": "2022-09-28",
}
```

```
"CYXSSn": "6",
"CYJSSn": "3",
"SJCJSJ": "2022-11-18 05:57:29"
}
```

加密后请求体参数示例:

```
{"cipherText":
"4c8fdbbbccf7f562515c525315dce16c7e6ba56b3b0e814cf7856e6fb5e2edd23
d88aeelcb11c460f2bfd0409d7b0c59e4c6743c3d46e65092907a87876d44f6f
c7d8272ee086d9127882303e05968d15ea779ba1000036854ce2894743a22f8e4
5e3fa5dcc946000f3dcf8cf018c990706c7ffd72544959e634ae0ee0aa8c52ea4
6fc604a47afcad9bc8ff1e835fcf50e81600e33f0ec2648169c542b3ca436017b
092164af06e8d67e8679695b9fa1e563731c53cee7ab39afb0cc5dfc4eeef5472
e3be5750659aacfc85ecd0d3678438a530b72773d54de1b4985956957yyy88828
3ed5235dd58f4e5f631c81eb05493c1b1f47b8be1d056f384f5103c7ba11fd5ac
8f6eebf2ad34afd26b2840ff6e085dbc947c79dc9c4733c2c3ff0169c04"}
```

返回示例:

```
{
  "returnCode": "200",
  "returnMessage": "提交成功, 数据处理中, 请 30 分钟后请调用增量查询结果
接口, 进行查看本次推送结果",
  "returnData": {
    "size": 1,
    "time": "2022-04-15T09:48:55.938"
  },
  "returnCount": null
}
```

4. 数据上报结果查询接口

接口描述: 按数据上报时间段查询数据上报结果

请求URL:

```
http://ip:port/prod-
api4/api/web/convergence/collect/interface/getResultsByTime
```

请求方式: POST (参数在表单(x-www-form-urlencoded)中)

请求头:

参数名	必选	类型	说明
Authorization	是	string	令牌
Content-Type	是	application/json	请求类型
app_key	是	string	应用授权校验码, 通过调用用户授权

			接口获取
timestamps	是	string	接口调用时客户端系统时间戳，单位：毫秒。参数值与国家授时中心标准时间戳比对差值不能大于 30s
random	是	string	接口调用时系统随机码，可自行生成
sign	是	string	接口签名，通过 BODY 中加密前上报数据 JSON 字符串+timestamps+random 进行 SM3 加密生成

请求体加密参数：

参数名	必选	类型	说明
startTime	是	string	开始时间 (YYYY-mm-dd HH:mm:ss)
endTime	是	string	结束时间 (YYYY-mm-dd HH:mm:ss)
page	是	int	页数
limit	是	int	每页大小
tableName	是	string	模糊匹配表名（空值查询所有）

请求体加密后参数：

参数名	必选	类型	说明
cipherText	是	string	值为密文

请求体加密参数示例：

```
{
  "startTime": "2022-04-15 09:48:00",
  "endTime": "2022-04-15 09:49:00",
  "page": "3",
  "limit": "3",
  "tableName": "o"
}
```

加密后请求体参数示例：

```
{"cipherText":
"4c8fdbbccf7f562515c525315dce16c7e6ba56b3b0e814cf7856e6fb5e2edd23
d88aeelcb11c460f2bfd0409d7b0c59e4c6743c3d46e65092907a87876d44f6f
c7d8272ee086d9127882303e05968d15ea779ba1000036854ce2894743a22f8e4
5e3fa5dcc946000f3dcf8cf018c990706c7ffd72544959e634ae0ee0aa8c52ea4
6fc604a47afcad9bc8ff1e835fcf50e81600e33f0ec2648169c542b3ca436017b
092164af06e8d67e8679695b9fa1e563731c53cee7ab39afb0cc5dfc4eeef5472
e3be5750659aacfc85ecd0d3678438a530b72773d54de1b4985956957yyy88828
3ed5235dd58f4e5f631c81eb05493c1b1f47b8be1d056f384f5103c7ba11fd5ac
8f6eebf2ad34afd26b2840ff6e085dbc947c79dc9c4733c2c3ff0169c04"}
```

关键返回参数说明：

参数名	说明
-----	----

timestamps	接受时间与每次调用增量采集接口返回 time 时间对应
sysName	调用用户名
tableName	表名
collectNum	收到数据数量
storageNum	入库数量
remark	异常信息

返回示例：

```
{
  "returnCode": "200",
  "returnMessage": "成功",
  "returnData": [
    {
      "id": 9,
      "type": "IncrSync",
      "sysName": "sa",
      "timestamps": "2022-04-15T09:48:55.348",
      "tableName": "ods_gzzyqksj",
      "collectNum": 1,
      "storageNum": 0,
      "remark": "数据字段数量与数据对象字段数量不符，或者主键为
null, 请检查数据...."
    }
  ],
  "returnCount": null
}
```