上海理工大学

智慧校园数据治理平台项目

信息化办公室

二〇二六年十一月

目录

1. 总体技术要求 1
1.1. 建设原则 1
1.2. 技术架构要求 1
1.3. 分布式关系数据库要求 2
1.4. 分布式非结构化（对象型）存储要求 3
1.5. 平台性能与稳定性要求 3
1.6. 集群与基础设施管理要求 4
1.7. 国产化适配要求 5
2. 建设内容 5
2.1. 数据治理平台 5
2.1.1. 数据管理平台 5
2.1.2. 数据开发平台 12
2.1.3. API 管理平台 14
2.1.4. 数据服务平台 16
2.2. 数据治理服务 18
2.2.1. 数据规划服务 18
2.2.2. 数据治理服务 19
2.2.3. 分类分级服务 19
2.2.4. 指标体系与数仓建设服务 20
2.2.5. 数据迁移服务要求 20
2.3. 数据应用 21

2.3.1. 一表通应用 21
2.3.2. AI 智能填报 23
3. 实施与运维要求 24
3.1. 项目实施要求 24
3.2. 培训要求 25
3.3. 运维服务要求 25
3.4. 安全与合规要求 25
3.5. 交付物要求 26
3.6. 团队要求 26
3.7. 其他 27

1. 总体技术要求

1.1. 建设原则

本项目围绕高校数据资源"聚、治、管、用"全生命周期，构建一体化数据治理与数据服务体系，遵循以下原则：

- **统一性。** 数据采集、治理、管理、服务、安全、应用等能力构建于一体化平台之上，各功能模块共享统一的元数据底座、权限体系与门户入口，不得以多个独立子系统拼接方式实现。

- **标准化。** 遵循国家教育行业信息标准，支持标准的导入、检测、发布、版本管理与落地对标，确保全校数据口径一致、命名规范统一、代码集可互换。

- **开放性。** 对外提供统一数据服务接口与开放门户，支持数据资产开放目录自助检索与申请，支持第三方系统以标准协议接入与消费数据。

- **安全性。** 对数据采集、传输、存储、使用、共享、销毁全过程实施分级分类管控，落实最小必要授权，保障师生个人信息与重要数据安全。

- **可扩展性。** 采用模块化架构，具备横向扩展能力，支持数据源类型、数据规模、并发访问量持续增长，支持组件自定义开发与动态加载。

1.2. 技术架构要求

1. 平台采用模块化架构设计、B/S 结构，开发语言为 Java，遵循 J2EE 技术规范；支持 Edge、Chrome、Firefox、Safari、360 等主流浏览器访问，具备移动端适配能力。

2. 支持云原生技术一键部署，提供高效易用的一键部署工具，可在工具内完成应用环境配置、私有镜像维护、服务依赖配置编排、应用升级与启动等工作。

3. 采用湖仓一体化数据架构，同时支持离线（批量）与实时（流式）两种数据处理模式；数据分层至少包含贴源层、标准层、应用层，支持扩展明细层、维度层、汇总层与历史层。各分层统一存储与元数据管理，支持批流一体的数据加工与流转，离线与实时任务共享分层模型和核心计算逻辑，避免同等逻辑重复建设；

4. 平台具备对海量数据全生命周期的分布式存储、计算、检索与开发能力；底座分布式组件支持在线安装部署与自动化运维，覆盖分布式存储、分布式协调、分布式消息、离线数据仓库、实时计算引擎、列式数据存储、权限控制框架、高性能搜索引擎、监控分析套件、分析型数据库、资源调度与管理、分布式计算等能力，支持组件不同版本的在线管控。

5. 平台应提供完整的技术架构说明，清晰描述业务架构、逻辑架构、技术架构、物理架构与数据架构，说明各架构分层的技术特点、选型依据，以及数据底座分层架构与数据中台、数据应用之间的逻辑关系。

1.3. 分布式关系数据库要求

提供全文检索、空间检索、时空检索、时序检索等数据引擎，支持高效的全文检索、空间检索与时序检索；支持关系型数据、JSON、XML 等数据存储及 NoSQL 数据库的外部数据包装器，实现 NoSQL 与 SQL 的有机整合；具备并行分布式查询引擎进行快速分析查询，支持列式存储压缩分区，兼容 SQL:2016、SQL:2011、SQL:2008、SQL:2006、SQL:2003、SQL:1999 及 SQL-92 标准；支持 OLTP 与 OLAP 并存（HTAP）模式、PL/pgSQL 存储过程语言与复杂 SQL 并行查询能力；支持自定义备份策略，可按备份集、时间点克隆实例并恢复到任意时间点以找回误删数据；支持并行计算，可利用高性能查询优化器自动根据 SQL 代价启动并行计算，覆盖表扫描、索引扫描、SQL 过滤、JOIN、分区表、聚合等复杂查询；支持图像相似度计算、高维向量计算、文本相似度计算、高维向量索引等扩展插件（满足图像

识别、人脸识别等场景）及树形结构、位图运算、高级位图等扩展插件（满足用户画像、实时推荐等场景）。

1.4. 分布式非结构化（对象型）存储要求

对象存储支持 Java、Python、Go 等多种语言 SDK；提供可视化对象数据存储管理平台，支持对象数据盘点（按时间过滤查看看板，查看对象数据桶及对象指标、服务在线指标、驱动在线指标、容量指标、网络指标，以时间维度查看存储趋势分析、对象存储大小分配分布分析、节点内存与 CPU 使用趋势分析）与桶及对象管理（列表查看已建桶详情与桶内对象列表，支持对象数据对外共享、下载、上传、删除，支持桶标签设置、对象复制规则、生命周期规则、访问权限规则设置）。

1.5. 平台性能与稳定性要求

1. 系统架构设计合理，能够稳定支撑日常办公、教务管理、业务流程、审批流转等校园事务工作，不出现因架构缺陷导致的频繁重启、异常闪退或不明报错，在校园高并发、多业务、长周期运行场景下保持服务不间断运行。

2. 数据同步性能。 系统应具备跨数据库、跨数据表的数据交换能力，交换过程中每秒处理数据量应大于 10 万条。测试以 1 亿条数据为样本，在源端与目标端的数据库表之间执行数据交换并统计耗时，每秒处理需满足“大于 10 万条/秒”的性能效率要求。

3. 数据查询性能。 数据总量为 2000 万时，1000 用户并发条件下平均获取数据时间小于 1 秒，且事务成功率不低于 95%。2000 用户并发条件下平均获取数据时间小于 1.5 秒，且事务成功率不低于 95%。

4. API 共享规模承载。 单节点应具备大规模共享 API 服务的创建与托管能力，能够支撑高并发条件下的持续调用，并在较大数据量列表查询场景下保持秒级响应。应支持创建不少于 500 个共享 API 服务,并可支撑日均 50 万次以上的接口调用。在对万条量级数据进行 API 查询时，接口响应耗时应控制在秒级，即不大于 1 秒。

5. 平台支持对数据集成开发的最大并发任务数、批任务并发数、单任务排队数进行设置；当调度任务达到最大并发任务数时，数据集成任务应能够自动排队。平台应支持横向扩展（增加节点）与纵向扩展（增加单节点资源），支持在线扩容，集群各组件支持高可用部署，年可用率不低于 99.9%，并提供完善的容错、降级与熔断机制。

1.6. 集群与基础设施管理要求

支持实时湖仓集群配置管理，包含主机、数据库和应用的新增、维护与删除；支持主机运行性能告警策略（磁盘、内存、CPU 使用阈值）、数据库运行状态告警策略（连接数、磁盘存储上限阈值）、应用连接状态告警策略及告警消息接收人自定义设置。提供实时湖仓集群预警消息展示，覆盖批流引擎、湖仓一体、文件存储、消息队列、指标引擎、搜索引擎、调度引擎、API 服务等组件的告警消息接收与查看。支持集群服务器运行状态监控（主机磁盘存储、内存使用、CPU 使用、I/O 读写等性能监控详情，服务器监控列表与主机详情查看）；支持数据库状态实时监测（数据库锁表异常与会话异常分析监测、异常解锁或终止、锁表与慢 SQL 分析、待监测数据库增删改查与告警设置）；支持文件存储情况监测（存储桶、存储对象、存储服务、存储驱动、存储容量扩展与网络特性监测，文件数据使用量增长、对象大小分布、API 数据接收率与发送率实时监测）；支持主机与数据库监测项配置（告警开关、存储/CPU/内存阈值、异常告警开关与接收人设置）。

1.7. 国产化适配要求

从信息安全与自主可控角度出发，系统建设应满足自主可控要求，能够在国产 CPU、国产操作系统、国产数据库、国产中间件等国产环境下平稳运行，并完成兼容性适配验证。适配范围应覆盖：主流国产芯片平台（包括但不限于龙芯、鲲鹏、飞腾等系列处理器架构）；主流国产操作系统（包括但不限于麒麟、统信、欧拉等发行版本）；主流国产数据库（包括但不限于达梦、人大金仓、神通、海量数据、GaussDB 等产品）；主流国产中间件（应用服务器、消息中间件等类别）。平台应支持在国产环境下完成部署、数据采集、数据治理、数据服务等核心流程的完整运行，并提供相应的兼容性适配证明。

2. 建设内容

2.1. 数据治理平台

平台须覆盖数据治理、数据集成开发、API 管理、数据资产管理与指标管理，并配套实施工作台、运营看板与数据安全管理，提供工程化、运营化管理能力。

2.1.1. 数据管理平台

围绕元数据、数据标准、数据建模、数据中心、数据质量、数据资产六个方面提供治理能力，支撑采集对标、标准落标、建模入仓、质量整改、资产编目与开放运营的全链路过程。

2.1.1.1. 元数据管理

1. 元模型与视图。 须内置关系型、大数据、文件系统等元模型包及数据源、表、字段、视图、主键、外键、索引等元数据类型，表模型属性覆盖技术、业务与管理属性，可新增自定义元模型与属性并维护其分类、解释与启用状态；须按"数据源—目录—表—字段"逐级下钻查看四层元数据，支持关键字搜索，各层级按基本信息与业务、技术、管理属性分类展示并可

编辑；须支持四层元数据整体发版，形成定版并保留历史、自动更新版本号，保证治理成果可追溯。

2. 采集与表分析。 须支持数据源增删改与连接测试，可按类型以基本模式或 URL 模式登记并绑定业务系统；须支持适配器查看、新增与修改，内置覆盖主流与国产化数据库、NoSQL、Hadoop 体系、消息系统及文件系统的适配器；须支持采集任务增删改与启停，按调度策略定时或手动执行并可多选对象，留存采集日志，支持单表重新采集与手工录入；须支持业务库单表分析，填写中文名称、表类型、入湖入仓情况与分析原因，支持批量分析、一键入湖与对标结果导入导出。

3. 可视化对标与链路分析。 须提供拖拽式对标，将源表与标准表拖入画布连线建立字段级映射，可查看字段详情与智能推荐标准；代码对标支持源代码表与代码标准字段连线，并可将整体代码表按规则拆分后对标；须支持按需扩标，记录每表完成字段数、标记完成状态与加载上次结果续做；须以可视化图形查看任一层元数据的上下游链路，提供全链、影响与血缘分析，可按表级或字段级定位，支持追踪单张表的数据在链路中的变化并检索某条数据对比其在上下游表中的差异。

4. 版本、变更与检索。 须以卡片与表格展示历史版本及版本对比，呈现各层次、各类型元数据的数量差距；须支持以树结构只读查阅任一历史版本的目录与全部元数据，满足审计需要；须提供变更统计与变更详情，业务库变更可手动执行并标记完成；须支持全局检索，按中英文名查询，按类型、来源、版本过滤并保存常用条件；须按业务系统统计采集表数、表注释率、字段数与注释率、各类型表数量及入湖入仓情况。

2.1.1.2. 数据标准管理

1. 标准与代码标准。 须以可视化方式维护信息标准、规范信息项唯一性，经录入与微调形成校级标准，管理字段名称、注释、类型长度、是否为空、是否主键及扩展属性；须支持业

务域大纲图谱阅览与树菜单拖拽、排序、挂载，业务域细分为目录与数据表，数据表作为版本发布与建模的引用对象，字段可跨表拆分合并；须支持代码分类与代码表管理，以规律编码规定代码集分类，代码表作为代码项集合支撑数据清洗，代码项可跨表拆分合并。

2. 标准导入与检测。 须支持按表、字段、代码、代码项从 Excel 模板导入，经规则校验后依新增、修改、异常、重复结果修正重校再审核入库；须实现标准自检，除非空、唯一、一致性、主键等校验外支持自定义规则并汇总为异常报告；须支持新增自定义规则，每条具备专属编号与一致性判断条件；异常可直接修改内容或字段、表信息并标注已处理。

3. 版本管理。 须展示当前版本收录量并按业务域、部门统计，新增、修改、删除的标准须发布版本并审核通过后方可被数据源引用；须支持浏览器内预览版本白皮书 PDF 并查询已发布版本；须支持版本审核，通过后成为最新版本并供标准网站与建模使用；须支持发布并校准与生效版本的差异，支持全量与自定义导出（Word、PDF、Excel），后台同时仅运行一个导出任务。

4. 参考标准与责任清单。 须内置面向教育行业的国家标准模板，主要来源于教育行业标准（2012 年版），并提供可视化维护接口，支持导入、新增、修改、删除；须支持查看各部门责任清单，可在表级与字段级维护并与标准联动，支持导入导出。

2.1.1.3. 数据建模管理

1. 建模报告与分层。 须展示数仓分层、主题分类、模型总数、近 30 天发布模型数、维表与事实表数及近一年建模趋势；模型须按公共维表、数据仓库、数据集市三个分层组织，各层一级主题可增删改并可查看主题下模型表。

2. 建模与核验。 须支持各分层二级及以下主题的增删改，主题内模型表可新建、编辑、删除、移动、发布，可查看物理化后数据与版本对比；须支持经数据库引用新建公共维表、经引用标准新建数据仓库模型，经数据库引用、SQL、自定义新建数据集市模型，支持拖拽组合

三类模型连线建表并自动生成物理建表脚本；须支持核验库表结构一致性，不一致时可修正并查看脚本。

2.1.1.4. 数据中心管理

1. 数据报告。 须展示数据湖（ODS）、数据仓库（DW）、数据集市（ADS）三层的业务系统数、业务部门数、表入湖率、数据表数、数据项数、数据量与容量，并可按层级下钻查看下级分类趋势；数据仓库层按业务域组织为主题库，数据集市层按分析主题组织为专题库。

2. 分层查看与字段级权限。 须支持查看数据湖表数据与元数据及文件数据、数据仓库事实表与维表数据及元数据（历史拉链表可查询历史数据）、数据集市表数据与元数据，各层均支持自定义导出与字段显示配置；须结合加密、脱敏与安全等级将访问权限控制到字段级，字段安全等级高于操作人员时不可见，设置脱敏时仅展示脱敏后数据。

3. 即席查询与供需分析。 须支持以 SQL 查询器与自定义 SQL 查询三层数据，并支持引导式选择多表关联、字段与条件自动生成 SQL；须以图形化方式呈现部门与系统间的数据流向，可按部门或系统过滤供需并下钻明细，以表格呈现表级供需；须支持全校需求清单的查看、维护并关联责任清单。

2.1.1.5. 数据质量管理

1. 质量概览。 须概览三层最新质量情况、核验趋势、改进动态与规则异常率排名，并可下钻查看各层各系统、业务域、专题的表级与字段级质量；须支持查看各部门质量概览，在数据湖与数据仓库层级下展示各部门质量情况及全部质量问题表。

2. 质量核验。 须集中管理内置、通用与自定义三类规则，按完整性、规范性、一致性、准确性、时效性、唯一性六维度定义，表级以 SQL 表达式、字段级以正则、范围、SQL 核验；须支持配置核验方案，按调度策略定时核验所选表及字段规则并可指定告警规则；须支持定时

或手动执行方案生成质检报告，亦支持单次快速核验；须留存核验日志（表数、字段数、规则数、问题数、得分）并支持报告导出、告警查看与维度权重重算得分。

3. 质量报告与改进。 须提供数据湖、数据仓库、质量改进及专题库报告，展示各部门与各专题表的核验得分与详情，可通过配置调整核验对象、章节内容与生成周期；须提供质量改进渠道，展示校内提交的反馈单并可受理转为工单，已受理工单可派发至业务单位并全程监控状态；须以图表概览问题类型统计、问题表 TOP5、各部门工单情况与平均处理时长；须支持各部门问题清单的查看、维护与工单下发。

2.1.1.6. 数据资产管理

1. 资产概览与编目。 须统计资产总数、已上架数、资产类型、上架与全部资产数据量的时间变化及各部门、各系统资产数；须支持全校资产编目，可单个或从元数据批量新增、修改、删除，导出目录并批量设置来源部门、来源系统、共享与安全等级，支持自定义目录逐级细分；须内置数据表、文件两类资产及属性，资产类型可增删改并配置属性（排序、是否导出项、是否启用），配置结果作用于编目页面。

2. 确源确权。 为落实"一数一源"，须以图形化方式呈现各部门、各业务域确源进度；须支持线上或线下确源任务的批量增减、导入导出并下钻至表级、字段级查看；对有争议或无来源字段支持二次确认，并提供未确源与已确源清单的查看、导出与导入结果处理。

3. 资产开放与运营。 须支持未上架资产的批量上架申请、开放属性配置与详情查看，已下架资产的批量上架与下架，并对上下架申请审批；数据表类资产须提供 API、ETL、中间表、离线 EXCEL、在线查询、实时订阅六种开放服务类型，文件类资产须支持预览与下载；须支持查看并回复使用人员的需求、建议与意见及其对资产、标准的评价评分；须支持处理各部门提交的编目、变更、上下架申请并查看回复结果。

2.1.1.7. 治理实施工作台

1. 待办、工单与预警。 须支持质量反馈单、需求申请单、部门资产申请等事项的详情查看与检索，可处理或回复并自动或批量移入已办；须支持实施工单查看与办理，办理后移入已处理；须支持质量核验、集成流程、集成资源、API 超时、鉴权与频率告警的查看，按状态、等级、时间检索并将未读标记为已读。

2. 月报与工程进度。 须支持实施月报的预览、下载与删除，每月最后一天自动生成并可手动生成，内容涵盖治理总体概览、数据采集入湖、标准修订对标、数据入仓入集市、资产目录与开放、质量改进等章节，以表格呈现明细及环比；须内置工程进度监控表模板，支持在线多人协作与导出 Excel。

2.1.1.8. 运营看板

1. 中台、三清单（即需求、问题、责任清单）与元数据。 须展现数据从业务系统流向数据湖、数据仓库、数据集市的动线，以及治理系统数、部门数、总表数、总数据量、容量与各系统采集情况；须展示三清单总项数、已落实项数与落实率、各部门落实排名、责任清单重要等级分布与应用支撑情况；须展示数据源数、业务系统数、采集任务数与表、字段总数，四层元数据的表数、字段数与翻译率及入湖入仓分析。

2. 资产、集成与数仓。 须展示资产运营各环节，含总编目、已确源、已上架与申请资产数、总申请与使用次数、申请量趋势、各部门上架与申请情况、共享与安全等级统计；须展示今日流程成功率、流程总数、数据交换量与入湖、入仓、入集市量，各业务系统集成表数与集成率，流程实时动态与近一周运行告警；须展示数据湖（含历史库）各系统入湖表数、字段数与数据量，数据仓库各主题库与各专题库的表数、数据量与容量，历史库近半年逐月变化。

3. 质量与安全。 须在主图轮播质量提升各环节及指标，展示各部门质量得分排名、改进排名与得分趋势，各业务系统与各维度得分及规则异常率 TOP5；须展示今日高、中、低危告警次数、脱敏与加密情况、安全告警趋势、审计事件动态、安全与共享等级分布及备份情况。

2.1.1.9. 数据安全管理

1. 安全概览与脱敏。 须概览安全审计事件、脱敏、加密、安全等级、共享等级与备份等整体情况；须支持建立脱敏规则，提供掩码、替换、重排、取整四种算法，支持脱敏策略的新建与管理并一键绑定字段及其上下游，可对不同操作人员设置脱敏权限，实现页面动态脱敏。

2. 加密、水印与访问控制。 须支持数据加密存储与 API 传输加密，仅授权人员或系统可解密，内置 SM4、SM2、RSA 等算法；须支持数据水印以追踪验证来源与完整性，策略可应用于各模块并支持明水印与盲水印，水印内容（文字、行数、大小、场景）可配置，盲水印文件可解析定位泄露源头；须通过认证授权控制访问，以公开、秘密、机密等级为人员与数据（到字段级）赋级并遵循"高密低访"原则；须以不予共享、有条件共享、普遍共享三级控制外部开放并匹配审批流程。

支持零宽水印数据安全溯源能力，在用户查看或导出数据时，系统自动在数据中嵌入不可见的数字水印。一旦发生数据泄露，可通过水印提取功能精准追溯泄漏源头，定位泄露数据对应的部门、用户名及操作时间，有效提升数据安全管控水平。

3. 备份与审计。 须支持新建备份策略将数据库保存在备用服务器，以 EXP 或 EXPDB 方式备份三层数据，留存备份历史（任务名称、数据源类型、方式、时间与状态）并可下载备份文件；须对数据访问与使用进行监控审计，通过记录分析访问行为及时发现安全威胁与违规行为。

2.1.1.10. 系统管理

1. 基础配置。 须支持维护全校业务部门、业务系统与标签的增删改；须支持调度策略的新增、启用与被其他定时任务引用；须支持流程的新增、修改、删除与图形化节点配置（审批人、审批条件、追加申请条件），可监控流程实例与状态，并维护动态负责人分组。

2. 运行环境。 须支持部署后配置数仓与共享库数据源、消息、API 中台、存储加密算法、容器平台等信息，并以向导化方式完成旧平台到新平台的数据迁移；须支持动态参数配置（导出模板、保密协议、Flink 访问地址等）与系统名称、Logo、校名、页脚设置。

3. 共享库与日志。 须支持查看可访问中间库的账号信息、修改密码，查看以中间表形式开放的资产及 ETL 流程，并对操作人员授权、查看关联视图与中间表数据、脚本 SQL 及取消授权；须支持制度文件与网站链接的增删改排序并发布至资产服务平台，可查看系统全部操作日志。

2.1.2. 数据开发平台

数据资源开发平台面向多源异构业务数据（含业务系统数据、线下数据、非结构化数据），提供数据源注册、源端数据盘点识别、数据集成入湖（贴源层）、入仓（标准层）、入集市（应用主题指标层）、数据计算、数据集成编排开发和数据挖掘算法开发等能力。平台对各类数据源的接入数量不作限制，包括数据库、数据集成接口、消息接口、日志数据源等。平台应提供数据治理整体进度看板，以项目实施视角展示系统部署、数据调研、数据实时入湖、数据标准建设、数据对标、质量检测、数据应用模型构建等内容的进展，并展示数据源层、贴源层、标准层、应用层各分层的建设明细。平台应支持数据集成开发的最大并发任务数、批任务并发数、单任务排队数设置，调度任务达到上限时能自动排队；支持智能化一键治理，通过选择数据源即可一键运行治理各环节。

2.1.2.1. 数据集成

1. 数据源与离线同步。须登记全部数据源，覆盖主流与国产化数据库、NoSQL、Hadoop 体系、消息系统及文件系统，集中管理同步任务，离线同步支持 MySQL、Oracle、SQLServer、PostgreSQL、金仓、达梦等关系库与 MongoDB 等 NoSQL 库，具备增量同步与位点记录、目标表自动创建与结构变更同步、异构类型转换、异常率熔断、分片并发与限速、断点续传及向导式多表迁移能力。

2. 实时同步。 须支持上述主流数据库及 Kafka 等数据源，基于 CDC 捕获变更数据实时写入数据中心，首次全量后自动切换增量；须支持向导式整库同步，可设定是否自动建表及写入规则，支持任务上下线、手动执行与各表同步情况查看。

2.1.2.2. 数据开发

1. 流程调度与版本。 须提供全托管的开发流程调度，可管理全部流程并支持导入导出、批量上下线、历史版本回退与回收站恢复；须提供失败策略、执行策略、优先级与全局变量等运行策略；须与数据处理知识库对接，自动生成清洗转换脚本并完成字段映射。

2. 可视化开发与组件。 须以 DAG 图拖拽编排任务，内置入湖、入仓、交换（抽取与加载）、网络（API 集成与推送）、脚本（SQL、存储过程、Hive SQL、Spark SQL、Python）、清洗转换（值映射、去重、字符串与缺省值替换、字段拆分多行）及流程控制（子流程、条件判断、选择）组件：入湖可直接引用同步任务，入仓提供关系检查与血缘查看并生成历史拉链表，抽取支持表单与 SQL 读取，加载支持增量写入与变更标记。

3. 实时与集成组件。 须提供可视化实时开发界面，内置实时输入输出组件、实时入仓、Flink SQL、Flink jar 等流开发组件及字符串替换、值映射、自定义转换器等流处理组件；文件集成须支持非结构化文档归集，将 csv、txt、excel、xml、json 等半结构化数据解析入湖，提供 FTP、SFTP、HDFS 文件与任务管理；日志集成须内置网络设备、系统、安全、网络行

为等解析模板，支持实时采集、向导式配置与图形化统计；消息集成须支持 Topic 新增、删除与授权访问管理，并提供消费监控。

2.1.2.3. 运维监控

1. 流程监控。 须提供离线与实时流程监控大屏，总览运行良好率、计划与实运行数、失败流程与实例数、同步数据量及流程总数，并给出失败、高耗时、大数据量流程 TOP10；须可检索实例记录、查看日志并下钻至任务级详情，以日历方式展示执行与调度分布。

2. 资源与告警。 须监控集群及各组件主从节点的 CPU、内存与磁盘使用率；须以监控规则对任务异常告警，提供实时告警与定期简报，可配置告警等级与范围；须监控数据交换链路异常波动，支持涵盖增删改量、增长率与总记录数的多维策略，阈值触发自动推送并具备处置机制。

3. 自动运维能力。对软件运行日志进行精细化监控，从多来源统一收集与存储日志，智能分析日志级别，优先关注 error 日志，并准确定位错误行，呈现系统整体视图。时间溯源法可快速排查定位问题，解决异常报错与历史问题。按周期生成运营报告，支撑重点业务保障、年终盘点、硬件采购规划及软件质量检测，为教师汇报提供全范围运维数据。

2.1.3. API 管理平台

支持总览 API 中台各维度统计信息并以可视化方式展现，包括接口分布、接口发布趋势、最新发布接口概览、各接口情况数量统计、接口调用趋势、接口调用排行、接口预警记录与趋势、应用运行情况、开发者授权接口统计、最新消息等。

2.1.3.1. 接口管理与安全

1. 统计与发布审核。 须提供 API 中台统计大屏，可视化呈现接口分布、发布与调用趋势及排行、预警记录、应用运行与开发者授权情况；须按应用或分类展示接口列表与详情（名称、

路径、所属应用、版本、上架状态）并支持历史版本对比；须由系统管理员审核发布的接口，仅上架接口可被调用，支持下架、重新上架并查看下架原因。

2. 申请与授权。 须支持开发者提交申请单申请调用权限，单张可含多个接口并允许部分通过；须支持系统管理员在接口上架后主动授权，可对不同开发者授予同一接口不同的数据与字段权限并配置授权 IP，实现同一接口供多应用使用且输出不同数据；须支持开发者账号维护、密码与密钥重置，密钥唯一且凭密钥调用，可配置变量编写前后置脚本。

3. 访问控制与加密。 须支持 IP 白名单机制，作用范围可设为全部接口、应用级或接口级，密钥泄露时未授权 IP 无法访问，并可通过黑白名单禁用或放行指定地址；须基于 OAuth 2.0 保障调用与数据修改安全；须支持 RSA、AES、DES、3DES 等对称与非对称加密算法及 SM2、SM4 国密算法，并支持面向不同应用、控制到字段粒度的掩码、替换、重排、取整脱敏。

2.1.3.2. 接口开发与编排

1. 开发方式。 须支持基于全量数据中心数据源及 Oracle、MySQL、SQL Server 等异构数据源创建接口，提供界面化配置、在线代码开发、领域驱动模型生成等方式，兼容第三方介入、结构化配置、模型快速开发、Swagger 导入、源代码开发等类型，支持筛选、导入导出、批量发布下架、编辑、回滚、复制、删除与历史版本查看，并可下载开发指南及 Java、PHP、C#等调用 demo。

2. 可视化编排。 为提升复用率，须提供可视化建模工具，拖拽集成接口、条件、循环、脚本、出参配置等组件，将多个已有接口重组为新接口，平台内编排无需编码；须支持跨应用编排、权限控制、子流程嵌套、循环与路由分支，支持页面配置、脚本配置或引用响应参数等出参方式及暂存复制；须具备错误检查与在线测试，可基于用例定位问题环节；须支持脚本沉淀为规则模板，实现类型转换、加解密与补全。

2.1.3.3. 测试、监控与中心

1. 自动化测试与调度。 须支持测试计划的查看、新增、修改、删除、批量启停与立即执行，提供免编码的自动化测试，一键发起即按用例执行并实时反馈结果；须支持对多个接口进行周期或单次调度，调度异常时自动邮件通知。

2. 运行监控。 须支持独立部署应用的运行查看与重启；须支持预警规则配置，覆盖响应时间、返回数据与调用频率三类异常，提供性能与异常原因分析、接收人配置、邮件短信提醒及预警处理；须提供应用、接口、测试、授权、调用、开发者等多维统计，并依访问量弹性调整资源。

3. 接口与开发者中心。 须支持全部已发布接口的详情与调用情况查看，按应用、标签、发布时间、名称与路径检索；须展示应用接口与预警统计、操作历史，含 API、数据源、已发布与回收站接口数、开发者数、接口状态、调用 TOP10 与趋势，并可切换被授权的应用；须展示开发者基本信息（appid、密钥、授权接口、到期时间、IP 授权）与调用统计，列出已授权及申请中的接口并支持新增申请，可查看调用次数、成功与失败数及各耗时指标，支持复制密钥、历史版本与多维检索，并提供开发者指南与文档下载。

2.1.4. 数据服务平台

平台支持数据开放目录的自定义设置，可维护多套目录并按主题分类进行多层级创建；数据资产可按需进行合并、拆分等资源结构定义并挂载到资产目录下，可与数据开放市场集市关联并灵活进行展示发布授权，支持开放市场数据集的灵活定义配置并挂载到数据市场目录下。平台应预置至少四种数据门户风格供选择，支持自定义风格配置与首页布局。

2.1.4.1. 数据门户与标准

1. 数据门户。 须支持开放统计概览展示，可按关键字查询资产目录、信息集与代码集，按领域、部门、专题浏览最新与热门目录，并支持制度文件查看下载与链接跳转。

2. 信息集与代码集。 须以卡片形式展示全部信息集标准，按业务域、标准类型、评分与关键字过滤排序，支持直接下载或加入选数车、收藏夹批量下载，可查看详情、数据字段、关联模型及数据项引用情况并评价；代码集标准须支持按目录层次、类型、评分与关键字查询，同样具备批量获取、详情与引用查看及评价能力。

2.1.4.2. 资产目录与申请

1. **检索。** 须提供校级数据超市，集中呈现已上架资产，支持按目录树、来源部门、共享等级、资产类型、服务类型查询排序；

2. **申请。**须支持加入选数车、收藏夹批量申请或直接申请；资产详情须展示数据项、数据预览、质量与元数据分析、申请与调用记录、问题反馈及历史评价；申请可单独或批量发起，填写个人信息与用途、选择字段与服务类型，可暂存或提交审批，多表多文件选择同一服务类型时仅填写一次。

2.1.4.3. 部门资产中心

1. 概况与确权。 须展示本部门已确源、已上架资产数，资产审批单与质量工单数，供需情况及供给、申请 TOP10；须支持处理下发的线上确源确权任务，在字段级确认归属后提交或暂存。

2. 审批与运营。 须支持业务部门审核节点的申请单逐一或批量通过；须支持查询本部门已上架与未上架资产并发起编目、上下架与变更申请，查看全部申请及回复，并从资产与申请单两个视角掌握本部门对外申请与外部申请的概况明细；须支持接收质量工单并批量分配处理

人，查看下载治理前、治理后与改进报告，以及本部门数据湖原始数据与数据仓库标准化数据的表、数据与元数据。

3. 文件空间与系统。 须支持本部门私密与公共文件空间，公共文件可编目为资产并全校共享，文件与文件夹支持上传、下载、删除、移动、重命名及链接分享；须支持本部门业务系统与数据源的查看与维护，部门主管理员可维护其他部门资产管理员账号与安全等级。

2.1.4.4. 统计与个人中心

1. 开放统计。 须展示信息集与代码集、开放数据表、数据项、数据量、文件数、开放API、开放部门与服务应用数等统计，以及各部门开放与申请数量、各领域开放排行。

2. 个人中心。 须提供个人统计概览，含选数车与收藏、反馈、评价、审批、工单、申请及已授权资产数量；须支持本人申请任务的查询与单个、批量审批；须支持接收处理质量工单；须支持查看本人全部申请及状态流转，按状态撤销、重新、追加、删除或查看详情，并查看获批资产与服务数据；须支持授权 API 在线调试、密钥邮件下发与 demo 下载，查看本人反馈回复及对标准、资产的评价。

2.2. 数据治理服务

2.2.1. 数据规划服务

1. **现状调研与需求分析。** 对学校现有业务系统、数据资源、数据质量、共享交换现状开展全面调研，梳理数据家底，形成数据资源清单；结合学校发展战略与业务需求，明确数据治理的目标、范围与优先级，形成数据治理顶层设计方案。

2. **架构与标准规划。** 规划数据分层架构（贴源层、标准层、应用层）与数据流向；制定数据标准体系，包括数据元标准、代码标准、指标标准、业务术语标准等；制定数据治理管理制度、流程规范与组织保障机制。

3. **实施路径规划。** 制定分阶段实施路线图，明确各阶段建设目标、交付内容、里程碑与验收标准；编制数据治理实施规范与操作手册，保障治理工作的可复制与可持续。

2.2.2. 数据治理服务

1. **数据标准落地。** 依据数据标准规范，对业务系统数据字段进行标准对标与映射，建立业务字段与标准字段的对应关系；对不符合标准的数据进行清洗转换，实现数据标准化落地。

2. **数据质量提升。** 建立覆盖完整性、准确性、一致性、及时性、唯一性等维度的数据质量规则库;开展数据质量检查与问题清单编制;协助业务部门完成数据质量问题的整改与闭环;建立数据质量长效机制。

3. **数据资产盘点。** 对全校数据资源进行盘点，建立数据资产目录;明确数据资产的权属、责任部门与使用规范；形成数据资产台账，支撑数据资产的统一管理与运营。

4. **数据治理运营。** 建立数据治理运营机制，明确各角色的职责分工；制定数据治理考核指标，定期评估治理成效；开展数据治理培训与宣传，提升全员数据素养。

2.2.3. 分类分级服务

1. **分类分级标准制定。** 依据国家法律法规与行业规范，结合学校数据特点，制定数据分类分级标准，明确分类维度与分级规则。

2. **数据分类分级实施。** 对全校数据资源开展分类分级，形成数据分类分级清单；对敏感数据、重要数据予以标识，为差异化安全管控提供依据。

3. **分类分级动态管理。** 建立分类分级的动态调整机制，数据新增、变更时同步更新分类分级结果；支持分类分级结果的查询、统计与导出。

4. 数据智能分类分级。支持一键导入《教育系统数据安全分类分级指南》文件，系统自动解析其中的数据分类目录，形成标准化的分类依据。支持通过 AI 对全域数据表进行智能分

类与自动定级，无需人工逐表配置。数据分类由系统从上传的官方指南文件中解析得到，与指南的分类目录保持口径一致。分级（重要等级）划分为 L1-L5 五级，由 AI 基于官方指南中的数据分类目录、分级标准、数据表名语义信息、分类分级知识库综合分析后判定。

2.2.4. 指标体系与数仓建设服务

1. **指标体系设计。** 围绕学校核心业务领域，梳理业务过程与决策需求，建立分层分级的指标体系（如战略层、管理决策层、业务执行层）；明确各指标的定义、口径、计算逻辑、数据来源与责任部门。

2. **数据仓库建设。** 依据数据分层架构，完成贴源层、标准层、应用层的数仓建模与开发；构建维度模型与事实模型，支撑指标的快速计算与灵活分析。

3. **指标开发与落地。**依据指标口径完成指标的计算开发与数据加工；建立指标血缘关系，实现指标的可追溯；支持指标的发布、变更与下线管理。

4. **数据分析应用。** 基于指标体系开展数据分析与挖掘，形成分析报告与决策看板；支持专题分析、趋势分析与对比分析，为学校管理决策提供数据支撑。

2.2.5. 数据迁移服务要求

1. **迁移范围。** 中标方应完成原有数据共享平台及数据治理相关系统的数据、任务与配置迁移，包括原有数据源配置、数据集成任务、数据标准、质量规则、API 服务、数据资源目录等。完成学校现有平台上 2800 余个在用 ETL 任务迁移，涵盖上行采集业务系统 60 余个，下行推送业务系统 60 余个，完成学校现有 20 余个 FDI 增量任务迁移及 50 余个在用 api 接口迁移。

2. **迁移实施。** 应制定详细的迁移方案，明确迁移内容、迁移步骤、风险预案与回退方案、迁移工期；迁移过程应保证数据的一致性与完整性，迁移期间应尽量减少对业务的影响。

3. **迁移验证。** 迁移完成后应开展全面的验证测试，确保迁移后功能正常运行、数据准确一致；应提供迁移报告，记录迁移过程与验证结果。

4. **平滑切换。** 应制定新旧系统切换方案，支持分批切换与灰度切换，保证业务连续性；切换后应提供一段时间的并行运行与观察期，及时处理发现的问题。

2.3. 数据应用

2.3.1. 一表通应用

要求覆盖报表与填报引擎、应用管理、问卷中心、模板中心、消息管理、学习中心、在线协同与移动端对接。支撑各类表单的自动预填、智能核验与在线填报。要求基于大语言模型与结构化数据解析，聚焦表格填报与数据录入场景，构建"智能识别—自动回填—辅助校验"的全流程助手并对接数据中台。

2.3.1.1. 应用与表单

1. 支撑引擎。 报表引擎须提供报表制作、填报、查询、打印与导出，支持填报表与统计报表，填报表支持拖拽配置并自适应移动端；填报流程引擎须支持可视化配置，提供多节点审批、抄送、分支及单人签收、多人签收、转办、节点字段权限、退回与自动跳过等场景。

2. 应用与表单。 须展示本人创建及他人授权管理的应用，支持新建、编辑、删除、分享与上下架，按名称、状态查询排序，可设基本信息、导航呈现与消息推送方式；可创建流程表单、数据表单、协同表单与统计报表：流程表单提供在线表单与流程设计、人员授权绑定、绑定主题表取数与查看授权；数据表单支持行级录入审批、在线建表、引用与导入建表，权限到字段与人；协同表单支持多人同时在线填写、实时同步与版本回溯；统计报表提供复杂设计器与常用图表组件，可拖拽搭建并按人员、角色、群组授权。

3. 待办与监控。 采集类表单须提供数据管理页汇总明细，数据表单可开放编辑与删除；须支持发起单次或周期性待办、循环策略与按期提醒，提供完成监控与页面催办；应用监控须展示待办、表单数量、使用情况与各单位完成情况，支持检索与催办。

2.3.1.2. 问卷、考试与模板

1. 问卷。 须以卡片与列表展示本人创建的问卷，支持检索及创建、编辑、数据、分享、停止、克隆、删除；创建支持拖拽题目，题型覆盖填空、多选、单选、时间日期、数字、评分、滑块、附件、取数、描述与计分；配置含收集起止时间、份数上限、隐藏题号等基本项，一次提交、可撤回、匿名、重复校验等高级项，公开、密码、校内授权等权限项及消息通知；发布后生成地址与二维码，支持 PC 与移动端填报；须提供提交与分题目统计的图表呈现，支持下载分享、筛选查询与批量导出，并分类展示可填问卷与已填记录。

2. 考试与模板。 考试须支持拖拽或题库导入组卷，题型覆盖简答、判断、选词填空等并支持分数与解析设置，提供期限、时长、防作弊、权限与消息设置，发布后扫码参加；阅卷须客观题自动评分、主观题人工打分并统计总分，支持多阅卷人与试卷导出，并提供分题目统计、明细导出及题库维护；模板须分为填报表单、调查问卷、考试测评三类，支持场景分类、上下线、移动分类与一键引用预览。

2.3.1.3. 消息、学习、协同与移动端

1. 消息与学习。 消息管理须集中展示通知与待办，可设置待填写、待审核、退回、催办等消息的发送渠道（系统消息、学校即时通讯平台、短信、邮箱）、免打扰与非工作日不发送策略；学习中心须提供快速入门、分步视频教学与操作文档入口，支持手册按目录或关键字查询。

2. 协同与移动端。 在线协同须支持多人同时在线填表、实时互通、更改记录按版本回溯，区分校内与公开协同；移动端须提供填报、审批、报表查看与个人数据查询，支持通过学校即时通讯平台直接登录，首页展示待办入口、应用、常用应用与问卷，可操作上下架、分享与收藏并查看待填报、待审批表单；须提供不少于 50 个应用表单。

2.3.2. AI 智能填报

1.智能建表。 须支持用户以自然语言描述建表需求，或上传文件由 AI 解析业务意图与文件内容，自动推断表名、字段名称等信息，一键生成符合业务语义的表单及配套填报流程模板；须支持以自然语言对表单及字段进行新增、修改、删除，AI 生成结果允许人工直接调整。

2.智能填表。 面向线上表单填报场景，须支持上传 PDF、Word、图片等，由 AI 抽取文档信息并一键回填申报表单，覆盖职称晋级、项目申报等复杂材料申报场景；须支持对接数据中台数据源，按表单格式取出数据、整理后自动填入。

3.离线文件智能填表。 面向线下流转、多源手工汇总的填表场景，须支持上传 Word、Excel 等离线表单文件，由 AI 识别文件结构、字段语义与版式布局并定位待填项；须支持统计类表格按维度匹配并生成统计结果填充；须支持人工确认后一键导出。

4.智能审核分析。 须支持按表单开启智能审核识别分析能力，对填报人提交的数据在审核阶段自动开展智能分析与校验；须内置默认审核规则，并支持按业务场景调整审核要点；审核结果须给出问题定位与修改建议。

5.AI 问卷创建。 须支持用户以自然语言描述问卷需求，或上传图片、PDF、Word 等材料，由 AI 解析业务意图与文件内容，自动推断问卷主题、题目、选项、题型与逻辑，一键生成线上问卷；须支持以自然语言对问卷题目与选项进行新增、修改、删除。

6. 代码辅助与智能问答。 须支持代码辅助，可生成组件样式、提供行内建议、辅助编写 JSHeader 脚本与 SQL 语句并展示变更对比；须支持基于操作手册等内容的智能问答，并对表单填报内容进行智能总结与分析。

3. 实施与运维要求

3.1. 项目实施要求

1. **实施组织。** 中标方应组建专业的项目实施团队，明确项目经理、技术负责人及各专业工程师的职责分工；项目实施期间应保持团队稳定，核心人员变更须经采购人同意。

2. **实施计划。** 中标方应制定详细的项目实施计划，明确各阶段的任务、进度、交付物与里程碑；实施计划应经采购人确认后执行，实施过程中应定期汇报进度。

3. 驻场服务。中标方应在项目实施期内安排 1 名专职技术人员常驻采购人现场，负责需求对接、数据治理实施、问题处理及培训支持，并按采购人要求考勤。项目经理与技术负责人应在需求调研、数据迁移、上线切换、验收等关键阶段到现场履职。驻场人员应具备数据集成、数据标准、API 与数据服务等相应岗位的实施能力，进场前须经采购人确认，实施期间不得擅自更换，确需更换的应提前 10 个工作日书面报采购人同意并保证工作衔接。

4. **实施规范。** 项目实施应遵循统一的方法论与规范，保证实施过程的标准化；应提供完整的实施文档，包括需求说明、设计方案、实施方案、测试报告、操作手册等。

5. **质量管理。** 中标方应建立项目质量管理机制，对各阶段的交付物进行质量审核；应配合采购人开展阶段验收与终验，对验收发现的问题限期整改。

6. 项目进度。项目总工期为自合同签订之日起六个月达到上线试运行状态，一年内完成项目建设并通过验收。中标方应在合同签订后 10 个工作日内提交详细实施计划，并在实施期内每周提交上周进度报告；因中标方原因导致里程碑延期的，按合同约定承担违约责任。

3.2. 培训要求

1. **培训对象与内容。** 应针对不同角色开展分层培训，包括面向管理人员的理念与规划培训、面向业务人员的应用操作培训、面向技术人员的开发运维培训。

2. **培训方式。** 培训应包含集中授课、实操演练、案例讲解等多种方式；应提供培训教材、操作手册与视频教程，方便学员后续自学。

3.3. 运维服务要求

1. 质保服务。 中标方应提供免费运维三年。三年免费运维期内提供软件升级、故障修复与技术支持（包含数据治理所涉及的所有工作）；质保期自项目终验合格之日起计算。

2. **响应时效。** 应建立分级响应机制，对影响业务运行的重大故障应提供 7×24 小时响应；重大故障应在接到通知后 1 小时内响应，响应后 1 小时内给出故障分析报告，一般问题应在 4 小时内响应；应明确问题升级路径与处理时限。

3. **运维方式。** 应提供电话、邮件、远程、现场等多种支持方式；应建立运维服务台账，记录问题受理、处理与闭环情况；应定期提供运维分析报告。

4. **巡检与优化。** 应定期开展系统巡检，及时发现并消除隐患；应根据运行情况提供性能优化建议与实施支持。

3.4. 安全与合规要求

1. **数据安全。** 项目实施过程中应严格遵守数据安全相关规定，涉及敏感数据的操作应采取脱敏、加密等措施；实施人员应签署保密协议，项目结束后应归还或销毁相关数据。

2. **合规要求。** 系统建设应符合国家网络安全等级保护二级要求及数据安全相关法律法规与标准规范要求。

3. **知识产权。** 项目形成的定制化开发成果的知识产权归属按合同约定执行；中标方应保证所提供软件不侵犯第三方知识产权。

3.5. 交付物要求

1. **文档交付。** 应提供完整的项目文档，包括需求说明书、总体设计、详细设计、数据库设计、接口文档、测试报告、部署手册、操作手册、运维手册、培训材料等。

2. **源码交付。** 涉及定制化开发的部分，应按合同约定交付源代码及相关技术文档；应保证源代码的可读性与可维护性。

3. **知识转移。** 应向采购人进行完整的技术交底与知识转移，确保采购人具备独立运维与二次开发的能力。

3.6. 团队要求

1. 项目负责人具备信息系统项目管理师、系统架构设计师证书的优先考虑。需提供有效证书扫描件及开标前六个月任意一个月为其缴纳社保的证明材料。注：证书属于计算机技术与软件专业技术资格。

2. 提供不少于 8 人项目团队（不包括项目负责人），团队成员中具有信息系统项目管理师证书、系统集成项目管理工程师证书、软件设计师证书、注册信息安全专业人员证书的优先考虑；提供 1 名专职人员在实施期内全程驻场且该人员需具有至少 2 个类似数据治理或数据中台类项目实施经验的优先考虑。需提供开标前六个月任意一个月为上述人员缴纳社保的证明材料。注明：除 CISP 证书外的证书属于计算机技术与软件专业技术资格。

3.7. 其他

1.供应商具备数据管理（数据中心）类、数据开发（数据交换）类、API 管理（接口管理）类、数据资产（数据服务）类软件著作权证书的优先考虑。

2.供应商具有近三年类似项目业绩的优先考虑。