

西北大学服务类项目采购合同

根据《中华人民共和国政府采购法》、《中华人民共和国民法典》及有关法律
规定，遵循平等、自愿、公平和诚实信用的原则，西北大学（甲方）与南京迪塔
维数据技术有限公司（乙方）就西北大学网络和数据中心（系）购买的西北大学
数据治理升级服务采购项目（招标编号：GCZB2025-12-239-Q）经双方协商达成
如下合同条款：

一、服务名称

西北大学数据治理升级服务采购项目

二、服务内容、要求及标准

详见附件。

三、服务费用及支付方式

（一）服务费用：大写金额 肆拾玖万捌仟元整；小写金额 ¥498000.00。

（二）支付方式和要求：合同生效后，待服务期满并经甲方验收合格后，一
次性付清服务费用。

四、服务期限

自合同签订之日起 180 日

五、知识产权

乙方应保证所提供的服务或其任何一部分均不会侵犯任何第三方的专利权、
商标权或著作权。

乙方保证所提供的服务的所有权完全属于乙方且无任何抵押、查封等产权瑕
疵。如有产权瑕疵的，视为乙方违约。乙方应负担由此而产生的一切损失。

六、验收要求

（一）质量标准

乙方保证提供的服务质量应符合中华人民共和国相关标准及相应的技术规
范、本次采购相关文件中的全部相关要求及乙方相关服务标准及相应的技术规范
中之较高者。

（二）验收组织

甲方负责按规定组织验收工作并出具验收书，涉及向学校师生提供的公共服务项目，验收时须邀请服务对象参与并出具意见。

七、甲方的权利和义务

(一) 甲方有权对合同规定范围内乙方的服务行为进行监督和检查，拥有监管权。有权定期核对乙方提供服务所配备的人员数量。对甲方认为不合理的部分有权下达整改通知书，并要求乙方限期整改。

(二) 甲方有权依据双方签订的合同约定对乙方提供的服务进行定期考评。当考评结果未达到标准时，有权依据约定的数额延迟付款或扣除履约保证金。

(三) 负责检查监督乙方管理工作的实施及制度的执行情况。

(四) 根据本合同规定，按时向乙方支付应付服务费用。

(五) 国家法律、法规所规定由甲方承担的其它责任。

八、乙方的权利和义务

(一) 对本合同规定的委托服务范围内的项目享有管理权及服务义务。

(二) 根据本合同的规定向甲方收取相关服务费用，并有权在本项目管理范围内管理及合理使用。

(三) 及时向甲方通告本项目服务范围内有关服务的重大事项，及时配合处理投诉。

(四) 接受项目行业管理部门及政府有关部门的指导，接受甲方的监督。

(五) 国家法律、法规所规定由乙方承担的其它责任。

九、售后服务

(一) 乙方对合同服务的质量保证期为验收合格后之日起 12 个月。

(二) 根据甲方按检验标准自己检验或委托有资质的相关质检机构检验的检验结果，发现服务的质量或性能与政府采购合同不符；或者在质量保证期内，证实服务存在缺陷（包括潜在的缺陷等），甲方应尽快通知乙方。乙方在收到通知后 7 天内免费维修或更换有缺陷的部分。

(三) 如乙方在收到通知后，没有弥补缺陷，甲方可采取必要的补救措施，

但由此引发的风险和费用将由乙方承担。

十、履约保证金

乙方在与甲方签订合同前，须缴纳履约保证金；履约保证金金额为成交金额的5%。缴纳方式：银行转账、支票/汇票/本票、保函/保险；退还方式：服务期满验收合格后凭收据和验收单复印件无息退还。

十一、违约责任

(一) 合同生效后，甲乙双方应按合同规定认真履约。合同履约责任只涉及合同甲乙双方，不考虑第三方因素。

(二) 如因乙方工作人员在履行职务过程中的疏忽、失职、过错等故意或者过失原因给甲方造成损失或侵害，包括但不限于甲方本身的财产损失、由此而导致的甲方对任何第三方的法律责任等，乙方对此均应承担全部的赔偿责任。

(三) 如乙方提供的服务构成第三方知识产权侵权，由乙方承担全部责任。

(四) 合同履行过程中，甲方应积极配合乙方进行验收以及验收前的外围配套等工作。否则，因此导致不能按期验收时，不能追究乙方责任；正常情况下应在验收合格后30天内按规定向乙方付款，最长时间不能超过60天。否则，按照每日利率万分之五支付逾期利息。

(五) 除不可抗力原因外，如遇下列情况之一者，乙方所缴纳的合同履约、产品质量保证金甲方有权不予退还，作为对甲方的赔偿：(1) 合同签订后不能按合同时限要求提供服务；(2) 所提供服务和合同不符；(3) 不能按合同履行；(4) 项目验收不合格。

十二、不可抗力事件处理

(一) 在合同有效期内，任何一方因不可抗力事件导致不能履行合同，则合同履行期可延长，其延长期与不可抗力影响期相同。

(二) 不可抗力事件发生后，应立即通知对方，并寄送有关权威机构出具的证明。

(三) 不可抗力事件延续30天以上，双方应通过友好协商，确定是否继续履行合同。

十三、合同的终止

(一) 本合同因下列原因而终止：

- 1、本合同正常履行完毕；
- 2、合同双方协议终止本合同的履行；
- 3、不可抗力事件导致本合同无法履行或履行不必要；
- 4、符合本合同约定的其他终止合同的条款。

(二) 对本合同终止有过错的一方应赔偿另一方因合同终止而受到的损失。对合同终止双方均无过错的，则各自承担所受到的损失。

十四、其他事项

(一) 合同经双方签字盖章后生效。本合同一式柒份，甲方执伍份，乙方执壹份，招标公司执壹份，执行完毕后自行失效。

(二) 下述文件为本合同的一部分，并与本合同一起阅读和解释，且具有同等法律效力：

- 1、合同附件 1：服务内容、要求及标准；
- 2、招标文件；
- 3、投标文件；
- 4、中标通知书。

(三) 在本合同执行过程中，甲、乙双方协商签订的补充合同与原合同具有同等法律效力。

(四) 未尽事宜，双方协商解决。

(五) 合同签订地点：西安·西北大学

(六) 合同签订时间：2026年2月6日

甲 方

单位名称 (盖章): 西北大学

地 址:

法定代表人:

委托代理人: (签字) 侯小件

电话:

传真:

邮编:

乙 方 (全填)

单位名称 (盖章): 南京迪塔维数据技术有限公司

地址: 南京市雨花台区软件大道 180 号大数据产业基地 2 栋 206-208 室

法定代表人: (盖章/签字) 丁建

委托代理人: (签字) 丁建

电话: 025-58820386

传真: 025-52315529

邮编: 210012

开户银行及账号: 招商银行股份有限公司南京江宁支行 125905581110301

招标代理机构名称: 陕西国创招标有限公司

联系人: 任倩

联系电话: 17629071197

合同附件：1、服务内容，2、服务要求及标准：

1、服务内容

序号	服务内容	数量	计量单位	金额(元)	税率
1	软件开发服务-数据治理升级服务采购项目	1	项	498,000.00	6%

2、服务要求及标准

技术参数与性能指标
<p>一、采购标的需实现的功能：</p> <p>1、治理体系优化服务</p> <p>充分参考数据管理能力成熟度模型相关规定，将现有的统一数据治理平台、统一数据开放平台等归集为一个整体，以多视角的形式提供便捷一体化数据管理服务。在不同数据分层上提供引导式开发逻辑和考核标准，充分考虑参与数据治理的业务系统厂商、业务部门相关人员、信息化单位人员、数据治理厂商及学校领导、普通师生在数据治理体系中的角色定位，通过在线的标准化开发和数据治理引导式工作台的形式，以更人性化、符合自然语言逻辑的方式，大幅提升数据治理的标准化程度，大幅降低数据治理学习门槛和操作难度。</p> <p>2、数据底层重构服务</p> <p>为数据中台提供完整的底层数据处理服务，包括流批一体数据集成、用于湖仓建设的分布式数据湖仓、用于流式数据传输的 Kafka 集群等。同时提供分布式、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，提供完整的大数据 Hadoop 生态等技术组件，便于扩展数据存储、计算、调度、共享等延伸需求。从业务层面上进行数据分层存储重构，将学校的数据后台分为数据源层、贴源层数据湖(ODS)、数据仓库标准层(DWD)，配合历史数据存储的扩充，构成数据实体存储的标准架构方案。</p> <p>3、数据治理优化服务</p> <p>重构数据资产管理中心，以数据集成为起点，融合数据治理、开放及数据资产运营的全生命周期管理服务，以落地场景视角为出发点，为数据管理与资产运营角色提供专有化工具服务，同时在数据可“管”与可“用”的基础上构建一套更加切实落地的中台化数据运营体系，围绕数据的“存、管、服、用”展开具体服务的建设。主要包括：</p> <p>3.1 数据资源开发服务</p> <p>面向学校信息中心管理人员及数据建设实施人员，按照数据分层架构，提供在线的业务数据源识别管理、数据批量入湖管理、数据标准集成管理服务。</p> <p>3.2 数据资源管理服务</p> <p>面向学校信息中心管理人员及数据建设实施人员，按照数据中台能力划分，提供元数据管理、标准管理、数据管理、历史管理、质量管理、数据安全、数据开放、数据监控等服务。</p> <p>3.3 数据资源运营服务</p>

面向数据管理人员、全校业务部门、院系、各类校园应用建设厂商、甚至广大师生提供在线数据资产目录编制管理、数据资产管理、数据门户(数据资产展示、数据资产申请/审核、数据调研/认领 /仲裁/质量管理)等服务。

二、采购标的需实现的目标:

以数据治理升级为抓手,逐步完善提升数据集成能力、安全管理能力、数据治理及开放能力,提升数据的标准化、规范化程度。数据治理升级服务将以国家相关标准、数据治理成熟度模型、相关法律法规为准则,实现如下目标:

1. 规范化治理: 加强数据治理工作的规范化和自动化水平

解决以往数据治理依赖人工操作,有较大的随意性和不确定性的问题,以数据治理升级为抓手,建立健全、充分落实一数之源管理规范,充分补充数据资源类型,通过技术升级加强数据共享时效性,参照数据管理成熟度模型国家标准,建立起符合学校自身规范的数据治理制度和平台,动态提升数据资源质量。

2. 智能化治理: 降低治理实施成本,提升智能化手段

解决传统数据治理过程中需要大量耗费人工进行数据识别、数据对标、数据映射的问题,纳入更多数据治理智能化手段,提升“智治”水平,实现智能化数据识别、转义、分析、挖掘,从数据实施层面降低实施人员压力,从而将更多的精力投入到核心数据监管、质量提升层面上来。

3. 安全治理: 加强数据安全能力和水平

补充原有数据治理平台缺少的动态安全治理能力,充分参考《网络安全法》、《数据安全法》、《个人信息保护法》、《密码法》及《数据安全能力成熟度模型》和《个人信息安全规范》的相关规定,以数据为核心,重点围绕数据生命周期,从组织建设、制度流程、技术工具等方面,实现校级数据的安全治理和功能建设。

4. 提升数据应用支撑效能: 流批一体提升数据同步与计算效率

补充原有数据治理平台缺失的流式数据集成、数据实时同步能力,解决如学生缴费、一卡通流水实时监测等高时效场景下的数据实时同步问题。基于流批一体数据集成引擎,构建实时数据同步、秒级准实时数据计算平台,实现数据应用的快速响应,打破传统数据集成引擎计算的延迟壁垒,重构数据计算架构,以更高时效、实时更新的方式为师生用户提供高效数据服务。

三、服务内容及要求

(一) 治理体系优化服务

(1) 针对接入的数据库数量、数据集成接口数量、接入 API 源数量、接入消息接口数量、接入日志数据源数量均不得有任何限制。

(2) 能够基于数据标准开发体系提供详细的数据治理实施方案,包括:各类数据识别方式、各种数据类型的数据入湖(含实时数据入湖)方式、标准化数据入仓(含自动化数据入仓)的方式、原有数据中心迁移等过程的详细实施操作过程。

(3) 为了管理部门能够准确掌握数据治理进程,需提供数据治理整体进度看板:能够以类项目真实的实施视角,按照如系统部署、数据调研、数据实时入湖及进度、数据标准建设、数据对标情况、质量检测情况、数据应用模型构建等项目实施内容的进展情况。能够详细展示

各数据分层(数据源层、贴源层、标准层)的建设明细, 供应商提供详细的软件截图及功能描述。

(4)应能够支持平台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, 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。

(5)为了提高数据治理效率, 需支持智能化一键治理的功能, 通过选择数据源, 能够一键治理实现数据实施过程各个环节自动运行, 治理内容包含标准建设、质量排查等。

(二) 数据底层重构服务

(一) 整体要求

满足大数据一体化数据治理需求, 具有对海量数据全生命周期的分布式存储、计算、检索、开发 的能力, 能够满足多源异构数据的高效检索、快速分析的业务需求;

能够提供大数据底座整体技术架构图, 需要详细描述各架构分层的技术特点;

提供整体底座建设的整体规划思路, 并以高度归纳凝练的方式, 描述大数据底座方案核心特性;

(二) 大数据底座管理中心要求

大数据管理中心能够基于国产化硬件设备、芯片和操作系统进行部署, 满足湖仓一体化数据治理需求, 具体包含:

1、需内置不低于 20 个分布式大数据组件, 提供在线安装部署和自动化运维能力, 如 HDFS 分布式大数据存储、ZooKeeper 分布式协调系统、Kafka 分布式发布订阅消息系统、Hive 离线数据仓库、Flink 实时计算引擎、HBase 分布式列式数据存储、Ranger 权限控制框架、ElasticSearch 高性能搜索引擎、Grafana 监控分析套件、Doris 分析数据库、YARN 分布式资源调度与管理平台、Spark 分布式计算系统等等, 支持组件不同版本的在线管控。

2、提供不少于 8 个大数据组件在线集群安装、配置过程。

3、支持集群主机的在线管理, 支持在线查看主机状态、主机名、IP 地址、内存及磁盘使用情况、平均负载情况、机架及 CPU 架构情况以及主机角色。支持主机标签、机架分配、Agent 设置。

4、支持在线安装主机, 安装过程自动完成主机环境校验、Agent 分发等设置, 支持批量添加新的主机。

5、支持在线查看和编辑主机告警指标, 告警指标支持告警指标表达式、告警阈值、级别、通知人员、告警策略及处理建议等设置。

6、支持实时湖仓集群配置管理, 包含主机、数据库和应用新增、维护和删除。支持主机运行性能告警策略设置, 包含主机磁盘使用、内存使用和 CPU 使用阈值设置。支持数据库运行状态告警策略设置, 包含数据库连接数、磁盘存储上限阈值设置。支持应用连接状态告警策略设置。支持主机、数据库、应用告警消息接收人自定义设置。

7、提供实时湖仓集群预警消息展示, 包含批流引擎、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 服务告警消息接收和查看。

8、支持实时湖仓数据库、服务、资源运行状态的实时监控, 包含数据库集群、批流服务、搜索引擎、Kafaka、API 服务、Minlo 集群、调度引擎、指标引擎的运行状态和主机资源磁盘存储、资源内存使用、CPU 使用、I/O 读写等性能的监控。

9、支持数据库状态实时监测，包含数据库锁表异常和会话异常分析监测。支持对异常情况进行解锁或终止。

10、支持文件存储情况监测，包含 BUCKETS 存储桶、存储对象、存储服务、存储驱动、存储容量扩展和存储网络特性的监测。支持文件数据使用量增长情况、对象大小分布情况、API 数据接收率和发送率实时监测。

(三) 数据标准分层规划要求

1、详述数据分层建设理论，至少包含数据贴源层、标准层。

2、提供数据标准分层整体架构，重点描述数据底层分层架构与数据中台的逻辑关系。

(四) 分布式关系数据库技术要求

1、提供全文 / 空间 / 时空 / 时序等数据引擎，支持高效的全文检索、空间检索以及时序检索。

2、支持关系型数据、JSON、XML 等数据的存储，支持 NoSQL 数据库的外部数据包装器，实现 NoSQL+SQL 的有机整合。

3、具备并行分布式查询引擎进行快速分析查询，支持列式存储来压缩分区。兼容 SQL:2016, SQL:2011, SQL:2008, SQL:2006, SQL:2003, SQL:1999 以及 SQL-92 标准。

4、支持 OLTP+OLAP 并存 (HTAP) 模式，支持 PL/pgSQL 存储过程语言、复杂 SQL 并行查询能力。

5、支持自定义备份策略，可按备份集、时间点克隆实例，通过克隆实例恢复到任意时间点，找回误删数据。

6、支持并行计算，能够利用高性能查询优化器自动根据 SQL 代价启动并行计算。能够覆盖表扫描、索引扫描、SQL 过滤、JOIN、分区表、聚合等复杂查询。

7、支持 Imgsmlr (图像相似)、cube (高维向量)、Smlar (文本相似)、Pase (高维向量模块) 等插件，满足图像识别、人脸识别等场景需求。

8、支持 ltree (树模块)、varbitx (位图模块)、pg_roaringbitmap (高级位图模块) 等插件，满足用户画像、实时推荐等场景需求。

9、投标人需提供分布式数据库相关软件著作权。

(五) 分布式非结构化 (对象型) 数据库技术要求

1、对象存储支持 Java、Python 或 Go 等语言的 sdk 支持。

2、提供可视化对象数据存储管理平台，支持以下能力：

(1)对象数据盘点：支持按时间过滤盘看板；支持查看对象数据桶(Bucket)及对象(OBJCT)指标、服务(SERVERS)在线指标、驱动(DRIVES)在线指标、容量(CAPACITY)指标、网络(NETWORK)指标等，并支持以时间维度查看存储趋势分析、对象存储大小分配(OBJECT SIZE DISTRIBUTION)分布情况分析、节点内存使用趋势分析、节点 CPU 使用趋势分析等。

(2)桶及对象管理：支持以列表形式查看已建桶详情(名称、空间使用、对象数、创建时间等)；支持查看桶内对象列表，并支持对象数据对外共享、下载、新增(上传)、删除操作；支持桶标签设置、桶对象复制规则设置、生命周期规则设置、访问权限规则设置等。

(六) 数据集成引擎要求

1、支持各类数据集成组件的管理功能，包括 app(独立应用类组件)、source(源端数据读取

相关组件)、processor(数据加工处理类组件)、sink(目标端写入组件)、task(数据处理组合任务)类组件的统一管理;支持查看组件名称、类型、版本以及 URI 信息。能够下钻查看每个组件的详细信息包括组件基本属性,组件高级属性等。

2、提供完善的流批一体化(流式数据处理、批量数据处理)的数据处理引擎,能够统一管理数据采集、调度、处理任务,实现实时、批量、计算作业能力。

3、支持流处理任务和批处理任务的分类展示,其中流式任务支持查看任务详情、运行数据统计,如表更新数、处理总行数、更新总行数、新增总行数、删除总行数、表结构更新个数以及最后一次更新时间。

4、支持数据集成过程的高容错性,能够兼容数据源端各类结构变化(DDL)的情况,能够不停服自动调整并集成,降低数据源端的结构变更所引起的故障率。

5、引擎应内置各类包括 jdbc、spark、syslog、time(时序数据)、shel¹、mongodb、http、jupyter、kafka、datax 等数据处理技术。

6、为了能够满足未来数据集成场景不断增长和变化的数据源类型,流批一体引擎应该具备较强的可扩展性,支持用户自定义开发数据集成组件,支持从 http 地址导入组件、从属性文件导入组件。

7、支持以图形化界面的方式查看流处理任务的编排情况,包括源端组件,目标端组件等信息,支持集成任务版本回滚。

8、数据集成落地能力要求:

具备实时无延迟数据实时同步的真实落地案例。

具备秒级准实时同步的真实落地案例。

三、数据治理优化服务

基于大数据底座能够为学校提供实时、全面、稳定、功能强大的全域数据资源开发、管理和运营能力。高效打通数据孤岛,精细化管控数据资源、不断提升数据质量,促进学校数据共享、开放生态的建设。

(一) 数据资源开发服务

提供面向学校多源异构业务数据(含业务系统数据、线下数据、非结构化数据)提供数据开发能力,提供各类数据源注册、源端数据盘点识别、数据集成入湖(贴源层)、入仓(标准层)、数据计算、数据集成编排开发和数据挖掘算法开发服务。

1、标准化集成开发模块功能要求

提供一套标准的数据集成过程开发软件,用来规范化实施过程,实现标准的数据治理方法论落地,规避由于实施人员的主观因素造成的开发过程不规范问题。具体要求如下:

(1) 业务数据源接入

1) 支持数据库源、API 源、日志源、文件源、非结构化数据源、物联网设备等至少 6 种方式在线可视化连接配置,能够在线进行对接管理,需详细描述六种数据源的对接参数项;

2) 支持各类主流数据库及国产化数据库类型,包含:oracle、mysql、sqlserver、postgreSQL、华为高斯、人大金仓、达梦、南大通用、DB2、阿里、腾讯、浪潮开务、sybase、mongoDB、HIVE、hbase、ES、TGDB、impala 等;

3) 支持根据云知识库及智能语义化匹配等手段,自动完成数据源相关元数据识别工作;

- 4) 支持对数据资源识别情况的 KPI 量化考核, 通过各种量化指标如主键识别率、表中文注释、代码表识别率等, 自动评估数据资源开发过程的标准化程度;
- 5) 支持 API 数据类型的多层结构解析, 完成 API 数据的主子表拆分, 可动态定位返回参数的数据坐标, 并可对返回数据的嵌套关系进行子表数据的解析;
- 6) 支持文件对象的上传并存储至平台文件服务器, 支持对文件数据进行遍历解析;
- 7) 支持半结构化文件例如 Excel 文件的集成, 支持表格数据批量入库。在特殊情况下能够通过批注解析等操作, 支持单元格数据按照字符拆分成多字段自动入库采集;
- 8) 具备通过可视化配置的方式, 实现半结构化及非结构化数据集成的能力, 具体包括 API 数据源在线接入能力, 实现 post、get、delete、put、options 的 API 接口请求方式, 实现 Basic、Token 的 API 接口认证方式, 实现文件数据集成, 实现 file、hdfs、ftp、sftp、http、s3、webdav 协议的文件数据集成对接, 实现 text、json、yaml、xml、csv、avro、orc、parquet、excel 的主流文件类型读取。
- 9) 支持数据源的复制移动、批量启停等操作。

(2) 实时数据湖(贴源层)开发集成

- 1) 由于业务源头数据库存在数据缓存、延迟落库的情况, 需要支持业务源数据的时间戳补偿机制, 保障数据同步完整, 同时当源端无时间戳字段时, 中台能够自动补充时间戳;
- 2) 数据集成过程应能够实现主键自动补齐、数据大小写忽略、时区设置、数据库时区设置、数据批处理数量设置、批处理线程数设置、任务运行内存设置、数据结构自动同步设置、大表分片、区间全量、熔断设置功能, 能够实现使用 JSON、XML、TXT 格式扩展源端和目标端相关集成控制扩展;
- 3) 支持代码数据的智能对标(根据资源识别自动形成对标精准化设置)、检测、差异转码, 需提供产品截图证明;
- 4) 支持将业务代码数据一键设置作为校标, 简化数据标准建设过程;
- 5) 支持业务数据在线转标, 提供业务数据和标准数据拖拽连线转换一键智能映射。
- 6) 支持数据同步过程中的断点续传, 支持在增量数据同步过程中, 如果关闭数据库服务, 在数据库服务重启后, 任务能够自动从最后一次成功处理的位置继续同步后续数据。
- 7) 支持增量模式的区间全量集成, 即在增量模式下, 根据集成场景的需要能够手动设置某个数据区间进行采集。
- 8) 为了能够提升贴源层明细数据的查询效率, 在数据集成过程中能够自动实现大表数据分片存储。
- 9) 数据同步过程应能够支持并兼容数据源结构变化同步, 包括新增表、新增字段、删除字段、非关键字段修改名称、修改类型后, 应能够实现目标数据库自动调整修复。

(3) 标准层开发集成

能够按照学校代码标准和数据资源目录标准, 建立数据标准层结构;

能够设置数仓元数据安全等级, 包括公开、受限、隐私、涉密等级别。

能够对数据汇聚和开发任务进行定时调度, 应能够设置分钟级和秒级精确度。

数据进入标准层的过程, 能够提供数据清洗、数据连接、数据过滤、数据去重、数据转换等常用数据治理组件。其中:

数据清洗组件能够支持自定义 SQL、多表连接、多表合并、数据去重、数据过滤、数据转换等功能；

数据连接组件能够支持多表关联设置；

数据合并组件支持输出字段列、合并列自定义设置；

数据过滤组件支持各类常用数据过滤条件设置；

数据去重组件支持字段值保留或者去除的设置；

数据转换组件支持代码差异一键纠正、一键转换的能力，并预置常用转换规则和通过代码方式二次开发转换规则。

数据熔断组件支持数据拦截、数据置空下发、数据停更拦截条件自定义设置；

支持标准数据集成日志的生成和查看。

支持实时监测数据模型属性（字段属性）变更，提供变动差异分析，支持根据差异情况自动生成模型同步的 SQL 脚本，支持一键修复差异（包括字段新增、删除、更改等变化）。

提供复合流集成任务能力，可将集成方案一致的多个流任务放置于同一个线程执行，有效地优化资源，降低流任务对服务器资源占用，提供集成任务的稳定性。

2、通用集成工具要求

在标准开发流程工具基础上，需要为我校提供一整套相对独立的集成工具，用以辅助日常数据开发，解决数据点到点数据集成、计算等需求。

(1) 自定义 ETL 工具

1) 支持用户对任意分层数据进行 ETL 任务开发；

2) 支持实时流（无延迟实时同步）、高频轮询（秒级数据同步）、低频调度（分钟及以上级数据同步）三种集成能力；

3) ETL 组件：组件库至少包含：SQL 查询组件（支持模板化 SQL）、数据连接组件、数据合并组件、数据过滤组件、数据去重组件、数据转换组件、数据映射组件（支持自动映射组件扩展）、文件映射组件、kafka 组件、http 组件、文件存储组件、代码对标组件、质量检测组件、脱敏加密组件、历史归档组件、消息预警组件，同时能够支持嵌入第三方引擎或脚本语言组件（spark、flink、sqoop、python、shell 等）；

4) 支持增量数据同步、全量数据同步两种数据同步模式；特殊场景下能够支持读取、写入的最大等待时间设置，避免因特殊情况造成的系统阻塞；

5) 增量数据同步模式至少支持时间戳、自增字段、时间戳加字段三种模式配置；时间戳模式下支持时间戳格式配置、时间补偿机制配置、初始时间位置设置；自增字段模式支持初始自增位置配置；

6) 自定义 ETL 能够支持各种类型数据源之间的双向同步，包括：数据库到数据库集成、非结构化到文件集成、非结构化到数据库集成、文本到 API 集成、API 到 API 集成、文本到消息集成、数据库到消息集成、消息到数据库集成、API 到消息集成、日志到消息集成、数据库到 API 集成、消息到消息集成等多种场景模式；

(2) 实时集成模块

1) 为专门解决时效性要求较高的同步场景，需支持独立的、可自由编排的实时流集成模块，实现各类点对点、自定义的数据实时同步；

- 2) 支持数据库源、API 源、日志源、文件源、非结构化数据源之间的双向实时同步；
- 3) 支持 CDC 实时流、轮询实时流两种实时数据源。
- 4) 支持在实时集成过程中对数据进行清洗转换，转换组件包含数据连接、数据合并、数据过滤、数据去重、数据转换、聚合分组、数据计算、数据输出等；
- 5) 支持图形化编排实时流集成、脚本开发实时集成，以及命令终端实时集成，需提供截图证明。
- 6) 计算组件支持对字段进行在线聚合计算；
- 7) 支持在同一个集成任务中拖拽多个不同数据库中的多张表进行关联、转换，关联后的数据能够实时集成至目标端。

(3) 批量集成

- 1) 支持创建批量集成任务，实现多个（不限数量）数据表一次性同步至目标数据库；
- 2) 批量集成支持多个不同数据库中的数据表，一次性创建集成任务并同步至目标数据库；
- 3) 支持批量集成过程的选项配置，包括不同数据库中字段类型的转换、源端数据过滤等；
- 4) 创建批量任务时能够支持表和字段的批量重命名。
- 5) 批量同步支持实时集成、高频轮询及批处理三种同步频率，支持全量、增量两种模式实现批量同步。

(4) 离线采集

- 1) 支持以离线采集任务的方式将线下业务办理过程中产生的 Excel、CSV 文件导入至数据中心；
- 2) 支持在创建采集任务过程，编写任务名称、注意事项等信息；
- 3) 支持离线采集任务的样式设计：单表或多行表格类型配置、字段显示与否与可编辑与否配置；多行表格支持数据操作配置（新增、更新、删除）；支持采集时间设置、开启采集提醒设置、采集人员范围设置等；支持采集任务二维码输出，方便采集人员采集数据。
- 4) 支持多个部门对同一个表格进行数据维护，对各部门的数据操作权限进行隔离。在同一表格中，同一部门下的多个用户可共同查看、维护本部门维护的数据，但不允许查看和操作归属其他部门的数据。
- 5) 支持离线文件采集，支持审批后入库与免审入库两种模式；
- 6) 离线文件导入时，支持全量导入（删除后新增）、仅新增、仅更新三种数据批量导入方式，更加精准的记录离线数据的变化情况。在数据导入前需要支持相关管理设置，如：代码字段名称是否转码、遇到某条数据导入失败时是否继续执行或终止回滚等；
- 7) 支持自动识别数据是否具有大字段，当有大字段时能够在线上上传导入文件填充至大字段，文件类型不限；

3、数据开发

- 1) 支持在线编辑 SQL 脚本，支持自定义函数、SQL 任务、存储过程、视图等在线创建、执行、编辑。
- 2) 支持在线编辑 shell 脚本，构建 shell 任务，辅助数据集成。
- 3) 支持在线编辑 python 脚本，自定义构建 python 任务，辅助数据集成。
- 4) SQL、shell、python 开发编辑器支持语法高亮、在线调试、历史执行情况查询。

5) 所有开发完成的脚本任务, 均能够在数据集成过程中作为集成组件自动展示, 集成任务中可以自由引用脚本任务。

4、算法模型开发

1) 提供在线可视化算法拖拽式编排子系统, 能够以可视化拖拽方式, 进行挖掘场景算法的编排构建;

2) 在线可视化算法拖拽式编排子系统需预置大量算法原子组件, 至少包含: 数据源、数据预处理(含缺失值处理、错误值处理、重复值处理、类型转换、数据过滤、异常检测、聚合、拆分等)、特征工程(特征编码、特征生成、归一化、标准化、主成份分析、卡方特征、重要性、特征选择)、统计分析(全表统计、相关性分析等)、机器学习(向量机、随机森林、决策树、逻辑回归、回归、预测等)、采样(随机采样、加权采样等);

3) 在线可视化算法拖拽式编排子系统支持编排画布放大缩小、全屏、鼠标拖拽等功能;

4) 在线可视化算法拖拽式编排子系统支持任意节点执行, 删除、查看执行结果及日志等操作;

5) 在线可视化算法拖拽式编排子系统支持算法调度配置;

6) 支持对算法的基本管理: 删除、添加、编辑、执行状态查看、启用控制、最近执行时间查看、手动执行等;

7) 支持组件的二次开发管理, 用户可自定义开发新组建;

8) 提供算法组件二次开发的 IDE 工作界面, 二次开发工作台应内置在平台内, 无需再跳转其他工作界面, 无需额外下载依赖包等操作。

9) 算法二次开发工作台支持 Python、R、Julia 等算法开发语言, 支持语法高亮和在线调试。

10) 投标方以三种或三种以上与学校业务场景相关的数据挖掘算法示例;

11) 中台内置算法开发, 应能够支持包括 Python 语言、Ruby 语言、C 语言、C++语言、Go 语言、Java 语言、Javascript 语言在线开发调试, 提供回归、聚类、文本分析、离群值分析、方差、偏度、时间序列预测算法。

5、物联网设备集成

1) 支持将多种设备、系统和数据整合到数据中台进行集中管理和实时监控, 可实现数据采集入库、集中存储、质量治理、安全加密和共享交互;

2) 支持 TCP、SSL、WS、WSS 多种安全协议接入;

3) 支持对设备网络组件配置、接入网关状态、设备状态进行实时检测, 保证设备可用性;

4) 支持对不同来源、格式数据进行解析处理, 将分散、异构数据整合为一致且可用的数据集, 为后续数据使用提供保障。

6、任务管理

支持任一层级的集成任务列表概览, 至少包含任务流向、执行计划、开始结束时间、日志查看、任务分析(数据流向拓扑, 数据明细、成功失败状态等)、数据对账、手动执行操作、删除任务、编辑维护任务、下线提醒等;

支持可视化拖拽方式对已建设集成任务进行组合编排;

支持常用组件, 如代码转标、质量检测、脱敏加密、历史归档、消息预警等;

支持第三方脚本组件拖拽并设置，必须包含 Flink、Python、Sqoop、Shell、Jupyter 等组件；

支持并行串行分支任务处理；

支持不小于 20 个节点的任务编排功能，需以实际案例详细描述任务场景及系统截图，并提供第三方实际使用客户的盖章证明材料。

7、运行监控

(1) 支持任务趋势分析，标注任务量核心指标状态趋势，能够以基准线、控制线标注任务指标线，趋势分析中要结合 CPU、内存运行负荷展开分析，需提供真实系统截图证明；

(2) 列表形式展示最近一周集成任务的执行统计，包括：集成任务平均执行时长排行、数据表平均读取量排行、集成任务平均读取量排行、数据表平均写入量排行、集成任务平均写入量排行等。需提供真实系统截图证明；

8、其他开发赋能工具

(1) sql 翻译工具

为解决异构数据库迁移等潜在需求，平台需支持 sql 转义组件能力，能够实现复杂 SQL 自动解析，在不同数据库之间实现一键转义，完成异构数据库数据的平滑迁移。

1) 支持 Oracle、PostgreSQL、MySQL、SQL Server、IBM DB2、达梦、高斯、人大金仓等十余种主流数据库复杂 SQL 语法精准解析，可识别 SQL 语句中各个组成部分，包括表名、字段名、关键字、函数、条件表达式、嵌套查询等；

2) 支持 SQL 语句自动转义能力，可自动将源端数据库 SQL 语句转换为目标数据库 SQL 语句，提供转换语义检查和 SQL 语句优化能力，确保转换后 SQL 语句逻辑正确性和执行效率，保证 SQL 语句在目标数据库中正确执行；

3) 支持不同数据库 SQL 方言差异处理，如特定的函数名称、数据类型、日期格式等，通过规则映射方式实现 SQL 方言的转换；

4) 支持 SQL 语句翻译结果对比，通过高亮比对方式展示语句差异，快速定位问题、验证翻译结果，省去繁琐的对标转标工作。

(2) 文档解析

1) 支持业务系统数据库字典表 (word、pdf) 在线自动转换生成格式化 Excel 数据的功能；

2) 文件转换支持表格标准模式、表格自由模式、表单批量模式三种转换形式：

表格标准模式，支持 word 文件直接生成标准数据识别调研 Excel 文件；

表格自由模式，支持在 word 中自定义批注信息，生成自定义表头的 Excel 文件；

表单批量模式，支持上传内涵多个同格式的 word 文件压缩包，将文件中的数据合并、自动转换为标准列表格式的 Excel 数据，例如，支持高校往年所有教职工年度考核表（复杂的中国式报表）批量转换到一个标准结构的 Excel 文件中，转换过程无需人员参与，提升数据采集能力和效率；

3) 支持 word 数据字典自动生成数据 DDL 结构；

4) 提供转换过程日志信息展示，便于记录转换过程信息；

(3) 数据比对

1) 能够通过界面设置，通过选择源头表和目标比对表的方式，创建数据比对任务；

2) 数据比对任务支持在线查看报告、下载报告;

3) 数据报告包括数据别比对情况(包括仅在源端存在的列、仅在目标端存在的列), 数据行比对情况(包括数据行数差异、仅在源端表中的数据 and 仅在目标端表中的数据), 数据值比对情况(包括同一主键数据在源端和目标端表中的差异, 例如: 值差异、精度差异、字段类型差异等);

(4) OCR 图文识别

1) 支持基于 OCR 技术的辅助数据录入能力, 提供从图片或剪贴板数据中监测并识别字符功能, 提供文字变形、倾斜、多字体、多字号等复杂场景下的文字检测识别功能;

2) 支持识别结果导出, 包含 PDF、WORD、TXT 等格式;

3) 支持对识别过程中错误字符及文本阈值进行精准设置, 支持错误字符处理方式设置, 可辅助精细化图文识别工作。

(5) 知识库查询

提供内容广泛、数据丰富的云端数据知识库(包含数据字段的定义、类型、取值范围基本信息和数据逻辑关系、业务规则、数据流动、关联等深层内容), 作为业务数据字典查询的辅助性工具。

(6) 数据水印提取

支持从带有水印的数据中, 精准、高效地提取出隐藏的水印信息, 可用于版权保护、数据溯源、安全审计等方面, 确保数据的真实性、完整性、安全性以及可追溯性。

(7) 系统迁移比对

系统迁移比对工具可用于辅助业务系统迁移, 支持对源系统与目标系统的数据表、字段、数据量等多维度进行精准比对, 快速找出差异, 清晰呈现迁移前后的模型变化情况, 助力用户高效识别问题。支持源系统和目标系统的一键迁移, 提供集成任务的自动化迁移, 可有效降低迁移风险, 提升迁移效率, 保障系统迁移顺利、稳定地完成。

(8) 数据全链路查询

数据全链路查询功能是基于数据血缘图谱提供对数据全生命周期进行全链路分析, 精准定位数据缺失问题, 可提升数据治理效率有效保障数据的完整性和一致性。支持查询任务的设置, 包含查询节点库、查询表、查询字段的设置。提供工具共享能力, 可将特定查询任务共享给业务部门进行数据查询。

(二) 数据资源管理服务

1、整体要求

数据治理管理所涉及的元数据管理、标准管理、数据管理、历史管理、数据安全、质量管理、监控中心、数据开放等模块必须为一体化平台, 不接受多个子系统拼凑。

2、服务要求

(1) 元数据管理

1) 支持批量数据库反向捕获元数据, 自动反向业务数据库中表及字段的中文注释信息, 且在业务数据库数据元素不健全的情况, 即无法获取完整元数据信息的时候, 厂商中台系统可依据自带的知识库体系(各类相关数据字典文档)对元数据信息进行中文语义的优化识别。

2) 支持元数据调度式采集。

- 3) 支持批量 Excel 文件导入、手动采集等数据模型线上管理、字段管理。
- 4) 支持元数据标签库的自定义建设，可对表、字段批量打标签。
- 5) 支持元数据地图查询，可按组织机构目录查看全校元数据资源情况。支持数据明细下钻，可查看每个数据源拓扑图。
- 6) 支持按照全业务域和实体数据库审计，统计各个元数据信息，如业务数据资源统计分析、业务元数据分类统计分析、业务元数据注释率排名、数据标准资源统计分析、主题应用资源统计分析等。

(2) 数据标准管理

- 1) 需预置国家教育部参照标准，并辅助学校编制执行的信息标准体系：基于国家标准、教育部标准、行业标准，结合学校实际情况编制和完善各项信息标准，给出信息分类编码规格说明书，建设形成一套符合学校自身信息化建设需要的管理信息化标准体系。
- 2) 支持数据标准、代码标准的日常在线管理，提供新增、维护、拆分、合并、变更、启停用等作业方式，支持批量化导入、导出操作。提供执行标准一键采用参照标准能力，可选择版本进行批量初始化导入。提供标准建设分析能力，可按指定标签对比采标标准和执行标准差异，快速了解目标建设情况。
- 3) 支持代码标准与业务标准的差异纠偏，可根据资源识别自动完成智能对标配置，同时提供在线勾选、一键智能映射配置操作。
- 4) 支持代码标准与业务标准差异捕获，提供在线对比结果查询、差异信息的统计分析、拖拽拉线转标、下载差异明细以及转码方式说明。
- 5) 支持代码标准与业务库非标准字段转标，提供现在勾选、一键智能映射、拖拽拉线转码等操作。
- 6) 支持标准版本管理。
- 7) 支持数据元素标准管理能力。
- 8) 支持数据确权中的数据预分配、数据认领流程管理，实现部门、业务信息员、信息中心多级审批流程。
- 9) 支持数据标准、代码标准表、字段列、数据行级别按业务源、部门、人员多维度权责管理。
- 10) 支持数据认领结果与权责的动态转换。

(4) 数据管理

- 1) 支持主数据管理能力，包含主数据目录管理、主数据全文检索、主数据创建、版本管理、备份、血缘查看；
- 2) 支持全域数据全文检索、高级检索；
- 3) 支持照片、文档数据等的批量采集，通过离线数据导入、导出操作完成数据采集，支持指定格式数据文件，包括 Excel、PDF、图片 ZIP 压缩包等，并提供 PDF、图片大字段的附件数据预览。
- 4) 支持标准资源简报、主数据报告输出。

(5) 数据历史管理

- 1) 支持完整历史数据切片留存和备份，确保历史数据的完整性和及时性。

- 2) 支持数据生命周期链，按天为颗粒度进行自动化数据切片链，具备充分自动化修复能力，能够在断电、网络异常等非正常情况下出现备份中断的历史“时间链”。
- 3) 支持历史数据黑名单和白名单两种归档机制，实现对历史归档数据的掌控配置。
- 4) 支持数据归档备份调度策略管理，提供备份任务执行的日志信息生成和查询，对非正常情况下备份失败的数据进行日志查询。
- 5) 支持监控全业务域数据范畴的活跃性，量化展示业务吞吐量；提供近一周、半个月、一个月、两个月、三个月、半年、一年、两年等任意时间周期内的数据变动情况统计报表、以及围绕数据动态时间轴事件。

(6) 数据质量管理

- 1) 平台必须预置常用的质量规则模板，包括但不限于：空值检测、重复检测、不在代码表检测、日期格式检测、证件号格式检查、数值范围检测、全角半角检测、长度不合规检测等，并提供可扩展的自定义规则模板管理能力。
- 2) 支持自定义质量检测项管理与配置，提供新增、修改、删除等维护扩充功能，可对数据进行业务检测项配置。
- 3) 支持质量规则模板、数据项检测规则的批量导入、导出迁移操作。
- 4) 支持业务源头一键全库规则设置，实现业务库质量自扫描。
- 5) 支持全面的数据质量自动化检测，可按业务生成质量分析看板、明细清单，提供在线统计、明细下钻、以及报告下载。
- 6) 支持按业务库与标准库质量评估等级和得分，可自动生成全面的质量画像、五维分析和质量趋势跟踪。支持数据质量绩效统计分析，可根据选择时间范围查看、导出数据质量修正情况。
- 7) 支持对质量问题数据进行自定义实时拦截下发的治理规则设置，控制脏数据流入业务，如“行不共享”、“单元格不共享”、“单元格数据置空共享”。
- 8) 支持对质量问题数据进行治理任务创建，可设置指定整改部门、整改人员对问题数据进行处理。
- 9) 支持对接数据应用，在线接收数据质量反馈、数据纠正消息，可进行线上诊断和消息推送治理作业流程，提供统计和报告输出下载。
- 10) 支持面向任何第三方提供质量反馈接口，保障数据治理体系的完全开放化和全校覆盖化。
- 11) 支持质量标签管理。
- 12) 支持按场景自定义报告输出。

(7) 数据安全治理

- 1) 支持数据 AES、SM2、SM4 加密交换功能，能够实现数据加解密；支持数据日期类型、INT 类型等非字符串类型字段的数据加解密；支持数据传输通道加密功能(SSL)。
- 2) 平台需预置常用的数据脱敏算法，为数据脱敏配置提供使用，包括但不限于手机号脱敏、身份证件号脱敏、姓氏脱敏、名字仿真等，且具备自定义灵活扩展数据安全加密、水印等安全算法规则。
- 3) 平台支持正则、算法、枚举等识别类型，预置身份证号、军官证号、手机、地理位置、

邮箱、护照识别等不低于 30 种敏感数据识别规则，建立识别规则库。

4) 支持密钥管理系统来实现自动加密，可使用唯一的对象密钥加密每个对象，该对象密钥由 KMS 管理的主密钥保护，实现数据库落库数据加密解密。

5) 可通过界面配置的方式，实现基于国密算法（SM2、SM4）的数据加解密集成。

6) 支持数据分级管理，能够自定义分级；支持基于敏感数据标签的自动数据分级。

7) 分类定级应支持基于敏感数据识别规则，对目标数据进行自动化打标及安全分级，实现数据自动分类定级。

8) 支持数据资产梳理、数据分类定级目录报批。

9) 支持数据异动及行为分析、预警、风险标识、追踪。

10) 提供 API 接口调用风险识别和管理能力，支持风险识别标签自定义设置，提供请求头、请求体、请求参数、URL、请求方法 5 中标签类别的设置，提供正则匹配、包含指定内容、未包含指定内容、等于指定内容 4 种识别规则表达式配置。

11) 支持 API 接口调用风险识别规则自定义配置，支持风险识别 IP 和应用白名单设置。

12) 支持 API 接口调用风险任务发布和处置。

13) 支持 API 接口调用风险预警管理，可及时通知管理人员进行风险处理，保障接口安全。

14) 提供数据仿真能力，帮助中台识别在的安全漏洞和风险。如数据泄露、非法访问等，确保数据在实际传输、存储、治理等环节中的安全性和系统稳定性。支持字段级仿真任务设置。支持加密转换、字符替换、位置调换、插入字符、随机字符串、固定字符串内置仿真规则设置，同时提供自定义仿真规则设置能力。

15) 应能够遮盖脱敏，包括保留前 n 后 m 字符、保留指定区间；应能够勾选使用哈希算法（MD5、SHA1、HA256、SHA512、HMAC、SM3）和加密脱敏（AES、SM4），应能够勾选使用国产 SM2、SM3 和 SM4 加密算法；应能够实现仿真脱敏，支持替换指定位置字符、指定位置调换、指定位置插入字符、随机字符串、固定字符串、数值加减乘除运算、随机数、日期前移后移指定天数、随机日期、时间前移后移指定秒数、随机时间规则。

(8) 数据开放管理

1) 支持 API 调用规则限制的自定义，提供 API 安全限制功能，如 API 接口的访问控制列表、访问次数、限制访问 IP 地址白名单、调用频率、调用时间区间、调用有效期等的自定义，可设置调用超限预警。支持向数据库视图申请进行授权账号密码、使用有效期、连接客户端的 IP 控制。支持审批者对申请数据的范围进行数据过滤设置。支持管理者选择部分字段进行授权、加密、审批通过或拒绝等操作，审批结果可在线查看。

2) 支持 API 接口、在线查询、文件下载、视图开放、数据订阅（含 ETL 推送/实时推送）等不少于 5 种类型服务构建。

3) 支持数据服务灵活配置，可在线进行图形界面单表编排可视化配置、多表拖拽连线方式构建数据开放服务，支持从可视化模式转为高级 SQL 模式构建服务，可灵活定义筛选字段进行封装数据接口；支持识别代码字段代码值转中文名，以及动态入参，支持数据过滤条件自定义，可根据业务需求对数据服务设置开放审批条件。

4) 支持数据服务申请审批流程自定义设置，提供无限级可视化编排申请流程的条件分支、审批节点、抄送节点，任一审批节点可面向指定人员、指定部门、指定群组、动态申请部

门、动态数据权责进行多部门联合审批设置，提供会签、或签的审批机制。

5) 支持第三方 API 的注册，包含 Restful、WSDL 两种方式的 API 接入，使得 API 通过统一的中台总线管理，保持原有接口的内容、格式输出，实现统一的鉴权管控；

6) 支持第三方 API 的二次编排封装管理，提供可视化和代码 2 种编排模式，可对 API 进行 Token 认证封装、输出参数格式调整、网络请求 (Http 请求、Post 请求、Get 请求、自定义验证) 设置、数据库操作 (数据查询、jdbc 组件、数据写入、自定义查询) 设置等业务编排改造操作，形成新的 API 进行开放使用。

7) 支持非结构化文件对象的开放授权。

(9) 数据监控

1) 提供集成任务数据自动对账能力。

2) 提供集成任务调度监控功能，能够根据调度方式、调度周期展示任务信息，包含实时流任务监控、低频调度任务监控；支持通过面积图展示实时流任务密集度，方便任务执行周期调整 (如横向时间轴展示低频调度任务条、任务密级度)；支持高密集度任务执行周期更改，防止因多任务执行消耗集成引擎内存而导致任务配额超限失败；支持任务卡片信息查看；支持任务执行时长预估，提供任务挂起预警提示，提前预知任务执行状态，减少任务失败可能性。

3) 提供数据库状态监测功能，能够根据不同数据库类型分页展示数据库状态监测列表；支持数据库状态监测详情查看，包含数据库名称、数据库类型、IP 地址或域名、端口、存储情况等。

4) 支持服务器监控列表展示，包含主机的运行状态、告警、CPU、存储、内存、IO 读写等；支持主机详情查看，包含主机基本信息、状态展示、启停历史清单。

5) 支持集成监测项配置，能够设置集成任务挂起阈值设置 (系统自带默认值，可手动更改)；支持全部、异常、挂起、数据更新异常集成任务告警设置；支持告警消息接收人设置，支持指定消息接收人、数据提供处负责人、数据接收处负责人设置。

6) 支持数据中台相关数据库的监测项配置：能够进行数据库锁表、慢 SQL 分析，能够提供中台相关待监测数据库的增、删、改、查，能够提供数据库告警设置功能，可定义是否开启告警，开启后可设置数据库连接数阈值、磁盘存储下限值、异常告警是否开启；支持告警消息接收人设置。

7) 支持数据中台相关主机监测项设置：提供主机告警设置功能，可定义是否开启告警，开启后可设置主机存储、CPU、内存阈值、异常告警是否开启；支持告警消息接收人设置。

8) 支持数据血缘分析，提供数据外部关联关系的图形化展现，动态展现数据集成的实际流向及来龙去脉，血缘图需定位字段级血缘信息展示。

9) 支持图形化数据追踪溯源，可进行字段级别的数据流出、流入路径，以及数据视图的真实来源解析。

10) 支持异常任务分析、挂起任务分析。

11) 支持任务依赖关系及先后调度时间的监控提醒。

12) 支持 API 集群监控，可检测主机状态、系统指标、CPU、内存、磁盘、网络等情况，及时识别和保护集群健康、性能和安全性等潜在问题。

(10) 人员数据治理

- 1) 支持人员多重身份类型标签在线管理，提供人员自动化识别能力。
- 2) 支持统一身份治理规则灵活设置，系统根据设定的治理规则可对身份管理进行检测，快速、高效定位人员信息质量问题。
- 3) 提供企业微信通讯录一键同步能力。

(三) 数据资源运营服务

1、整体要求

支持数据开放目录的自定义设置，可维护多套目录，按主题分类进行多层次目录创建。数据资产可按需进行合并、拆分等数据资源结构定义，并挂载到数据资产目录下。数据资产与数据开放市场集市可以进行关联，可对数据资产灵活进行展示发布授权。支持开放市场的数据集市灵活定义、配置，并挂载到数据市场目录下。

支持至少预置 4 种数据门户风格可供用户选择，并支持自定义风格配置、自定义首页布局。

能够提供英文版数据门户，支持用户在数据门户使用过程中英双语自主切换，需提供真实系统截图证明材料。

2、服务要求

(1) 运营管理

- 1) 支持数据资产、开放市场目录自定义，可维护多套目录，按主题分类进行多层次目录创建，包含新增、删除、修改、启停用操作。
- 2) 支持数据资产灵活配置，可对全业务域数据进行多表字段合并、单表字段拆分等按需数据资源结构定义，并挂载到数据资产目录下。
- 3) 支持数据资产与数据开放市场集市进行关联，并可对数据资产灵活进行展示发布授权。
- 4) 支持开放市场的数据集市灵活定义、配置，并挂载到数据市场目录下。

(2) 门户布局管理

- 1) 支持数据门户首页的自定义配置，可根据实际业务需求自由组合个性化内容，构建专有工作门户；
- 2) 提供 55+个性化组件，支持不同组件属性自定义设置；
- 3) 支持多样性门户定义，可根据不同用户角色授权不同首页。

(3) 数据门户

- 1) 支持数据治理成果、资产概况展示，提供按不同目录维度进行数据资源统计的功能，提供数据资产开放、申请及调用情况、采标情况的统计，可提供热门资产排名统计等。
- 2) 自动关联中台对于资产的管理权责、业务来源、安全定级、引用国标等标准，实现按主题、部门权责、业务来源、安全等级、国家标准的资产分类统计。
- 3) 支持按资产点赞量、浏览量、更新时间进行数据资产排序检索。
- 4) 支持信息标准公示网查询，包含数据标准、代码标准，提供标准编号、引用国标编号、字段数、公开类型、提供方式、提供状态、标准说明、字段列表、字段类型长度、代码内容等信息查阅。
- 5) 资产目录支持多种开放类型的卡片标签，可一站式批量申请使用，使得数据开发者能够快速获取数据，并进入数据审批流程，包含：API 接口、在线查询、文件下载、视图直连、

数据订阅。

6) 支持数据资产明细查询，如数据基本信息、资产所属分类、资产责任部门、资产来源部门、资产来源系统、数据来源方式、共享等级、申请使用次数、查询次数、反馈数量、数据字段信息、数据明细、数据血缘、数据质量、互动反馈等。

7) 支持审批通过的数据查询权限者进行数据资产的在线明细数据查询、下载。

8) 支持与数据安全水印设置的联动效应，被设置水印包含的数据展示明细加入水印。

9) 支持按自定义的开放市场目录进行数据集市的展示、检索；提供申请状态、开放类型、数据支撑来源等过滤条件的查询能力。

10) 支持对开放的数据进行查询预览，详细如开放数据的名称、描述、开放技术种类，开放的接口入参、出参（字段）、业务参数示例、业务返回示例以及调用的说明文档等。

11) 支持开发者在线发起开放数据的申请，可根据当前开发数据支持的开放类型进行“API 接口、在线查询、文件下载、视图直连、数据订阅”多类型一站式申请。

12) 支持以校级资产、部门资产两种视角查看资产目录清单、资产分析。

13) 支持全量化数据检索能力，可根据关键字进行业务元数据、主数据、非结构化数据等所有建设范畴内的内容信息匹配检索。

(4) 工作台

1) 支持面向部门及校级管理审批者的数据申请在线审批管理，包含 API 接口审批、文件下载审批、在线查询审批、视图直连审批、数据订阅审批在内的多种数据申请查询和在线审批。

2) 支持审批限制规则的自定义，提供安全限制功能，如数据调用的访问控制列表、访问次数、限制访问的 IP 地址白名单、调用频率、调用时间区间、调用有效期等的自定义，可设置调用超限预警。支持向数据库视图申请进行授权账号密码、使用有效期、连接客户端的 IP 控制。支持审批者按照字段级别选择字段授权、加密、对申请数据的范围进行数据过滤设置、审批通过或拒绝，审批结果可在线查看。

3) 支持管理者对申请进行回收权限或者修改授权，进行授权终止或授权矫正控制；申请者对于拒绝或终止授权的数据可重新发起申请流程。

4) 支持面向申请者的数据申请在线管理，包含 API 接口申请、文件下载申请、在线查询申请、视图直连申请数据订阅申请在内的多种数据申请查询。提供已申请的所有接口审批流程跟踪查询、撤销申请管理能力；申请者对于已撤销的数据可重新发起申请流程。

5) 支持面向审批者、申请者在线接口审批流程节点的流转状态流程图查询。

6) 提供开放接口数据血缘查看和数据反馈互动。

7) 提供应用转交功能，可将当前用户所属应用一键转交其他人员，方便应用后期维护，且不影响接口使用。

8) 支持业务数据在线调研任务接收，并完成线上资源调研识别表格登记工作。

9) 支持数据权责认领任务的接收以及流程流转，一级认领员可在线完善字段级认领或仲裁反馈，二级审批者可接收一级认领任务提交后流转，进行线上审批或仲裁反馈，审批可通过与管理平台权责进行实时同步，审批拒绝可退回一级认领员重新认领。

10) 支持数据质量治理任务的接收以及处理，可获取质量文档清单、质量数据明细，并在线

维护、跟踪处理进度，完成治理任务。

11) 提供按审批、任务的状态进行事项统计，支持申请审批者、任务处理者按待办事项的状态进行过滤。

12) 支持数据预约，可以让用户在线提交开放市场所不存在的数据资产申请。

13) 支持部门资产的管理，查看部门资产整体分析报告、质量报告、共享报告

14) 提供下行数据库转交功能，可将当前用户管理的数据库一键转交其他人员，方便后期维护。

四、性能及稳定性要求

(1) 实现在同一张原表存在多类型字段，且字段类型包含 BLOB、CLOB 等大字段情况下的实时数据同步，时间延迟不超过 1 秒。

(2) 数据开放共享：在数据总量为 2000 万时，2000 用户并发时平均获取数据时间小于 0.5 秒；3000 用户并发时平均获取数据时间小于 0.8 秒。

(3) 数据采集：在数据库表写入总数据量为 2000 万的数据耗时小于 30 秒。

(4) 数据资源门户申请管理：开放接口、数据查询、文件下载、视图直连、数据订阅申请管理支持不低于 1000 用户并发查询，平均响应时间小于 0.3 秒，事务成功率 100%。

五、国产化适配要求

从信息安全角度出发，系统建设应满足自主可控要求，能够在国产 CPU、操作系统、数据库等信创环境上平稳运行，需同时具备以下条件：

提供 2 份及以上国产芯片（如龙芯、鲲鹏、飞腾等）的兼容认证证书；

提供 2 份及以上国产操作系统（如麒麟、统信、欧拉等）的兼容认证证书；

提供 2 份及以上国产数据库（如达梦、腾讯、人大金仓、海量数据等）的兼容认证证书。

六、质保期和服务期要求

质保期：自验收合格后 1 年，质保期内包含版本升级。运维期：自验收合格后 3 年。

四、其他要求：

1. 售后响应时效：响应时效：即时响应（包括电话响应）；电话响应无法解决 24 小时内到达现场。修复时间 12 小时内解决；如 12 小时内无法修复，应提供相应解决方案。

已确认附件内容。

高垣

2020.2.6

