## 一期项目明细表

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **建设名称** | **建设内容** | **报价** |
| 1 | 数据集成 | ETL数据采集工具和数据采集、清洗，转换等全过程管理。 |  |
| 2 | 数据管理 | 全域数据库管理，数据标准管理，数据模型管理，元数据管理等。 |  |
| 3 | 数据共享 | 数据资源对外共享服务的功能板块，它包含数据开放管理、数据脱敏加密、数据安全等级管理。 |  |
| 4 | 数据门户 | 作为整个数据中心项目的前端，提供数据资源查看入口，管理各角色对数据资源的浏览、检索权限。 |  |
| 合计 |  |

**硬件资源需求**

| **序号** | **服务器用途** | **数量** | **CPU** | **内存** | **硬盘** | **报价** | **备注** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 数据库部署服务器 | 2 | 32核/台 | 128G/台 | 4.8T/台 |  | 物理机 |

## 1.2一期项目招标软件技术要求

|  |  |  |
| --- | --- | --- |
| **序号** | **功能项** | **详细参数** |
| **1** | **数据集成** | **1.1ETL数据采集工具**基于先进的技术架构进行开发，支持基于Web界面即可完成相关功能的配置，支持单机多线程模式完成ETL任务的运行，支持进行分布式部署。**1）ETL接口开发**支持创建普通、批量两种不同的ETL接口创建方式。支持各种选项配置，包括字符类型字段长度扩充倍数、Char类型是否转换成varchar类型、源表为空是否继续执行等；批量ETL接口的命名规则可根据学校业务需求进行自定义配置。支持基于图形界面配置完成普通ETL接口的开发，完整配置ETL接口的各项参数，可支持将物理表、视图或SQL脚本定义为源表，可采用拖拽式操作配置字段映射关系，可在图形界面下配置主键字段、配置代码映射关系、多表关联规则、数据过滤规则、添加时间戳、应用SQL函数等。支持通过自定义SQL语句进行复杂逻辑的数据处理。数据抽取方式需至少支持全量抽取和增量抽取两种类型。支持将一个接口任务切分成多个数据块执行，提高接口任务执行成功率；支持通过画布拖拽的方式进行源表和目标表的选择，支持按照表、视图、同义词、SQL查询的方式进行源表的选择，支持通过表名称或中文注释进行目标表的检索和选择。支持在图形化界面中配置代码映射关系。ETL执行过程中，自动引用代码映射规则，完成对目标表代码的标准化转换。该功能与自定义SQL语句中的数据处理逻辑联合使用、同步运行、同时生效。支持拉链表模式向目标数据对象同步数据，实现目标数据的内容只增不减，保存数据行级颗粒度的历史版本。支持多种数据采集，包含但不限于api接口数据，MySQL数据、Oracle数据、SQLServer数据、PostgreSQL数据等。**2）异常情况处理**针对数据内容可能出现的异常情况，要求ETL功能在数据同步过程具备一定的预防性容错机制。包括：当源表为空表时，或当源表数据量与前一次同步时相比变化量超出设定的阈值时，接口将自动停止执行，交由人工确认后再决定是否执行。接口任务运行过程中个别数据错误时支持自动跳过并继续执行（避免整个接口运行失败），在接口执行成功后支持查看同步失败数据，只需重新对有错误记录的数据进行抽取即可。数据库连接可能出现的异常情况，要求ETL接口执行时具备合理的重试机制，当数据库连接失败、网络通讯故障等情况下，通过合理重试尽量提高接口运行的成功率。**3）接口管理**具备丰富的ETL接口管理功能。可显示ETL接口的运行状态和结果。可展示ETL接口的等待、执行、队列、成功、失败和挂起等不同状态，用户可基于对状态进行检索，快速查找各种状态的接口。可显示ETL接口的配置和数据状态。可查看接口的配置信息、状态监控和运行历史。状态监控和运行历史信息须包括运行结果、处理速率、堆积数据情况，每一次运行时读取、新增、修改、删除的记录数等，可查看每一次的详细运行日志，可快速定位日志中需要关注的重要信息，以提高排错效率。可对ETL接口的配置信息进行批量导入导出，以便快速完成接口信息的备份与还原操作。须支持基于对现有的ETL接口的复制编辑生成新接口。管理员将接口删除后支持暂时存放至回收站，便于用户误删后的回复、弃用接口的重新启用及已删除接口的批量管理。可通过web界面对数据库存储过程、Shell脚本进行调用，可载入前期留存的存储过程/Shell脚本模板，快速创建新的存储过程/Shell脚本。支持批量完成接口的管理和运行控制，包括批量执行、批量终止执行、批量为接口设置分类、批量导出接口及批量删除接口。为便于查看和管理大量接口，支持自主进行接口分类管理，可按需将接口划分到不同的分类下。支持按照接口名称、接口类型、接口状态、源库、目标库、只显示错误记录数进行接口的筛选和查看。支持进行接口配置信息、接口监控信息和接口运行历史信息的查看，接口监控信息支持对接口成功率、平均处理速率、堆积数据情况进行查看，支持对独处记录数、新增记录数、修改记录数、删除记录数、错误记录数、接口日志进行查看；接口日志支持通过定位关键字如“error”、“warn”等进行详细错误日志的查看，支持错误日志的全屏查看和复制。**4）性能**数据采集性能要求单节点离线抽取速率不低于50Mb/s，最大平均速率不低于20W条/秒，实时抽取效率可实现秒级同步，实时抽取单节点最高可支撑至少20M的峰值流量。**5）任务调度**支持基于web界面的调度任务创建，为降低使用门槛，产品需支持通过拖拽的方式实现同步任务、异步任务、串行、并行、任务流的创建，可根据我校数据的业务属性创建不同的执行频率或周期，支持单次或者周期性任务创建；支持基于Web界面的画布功能快速进行调度任务的创建，包括批量导入调度任务（接口、存储过程、Shell脚本）、自动连接接口、一键格式化、调度任务保存（调度任务名称、调度任务分类、任务类型、执行频率、执行时间、任务有效期等信息的配置）；管理员将调度任务删除后支持暂时存放至回收站，便于用户误删后的回复、弃用任务的重新启用及已删除调度任务的批量管理功能。**6）告警**为便于管理人员第一时间掌握ETL接口故障，产品支持创建告警任务，对ETL接口的运行状态进行告警监控，可根据不同的业务需求设置告警的触发事件。支持查看告警任务，当任务中的接口满足触发事件时，至少可通过邮件对指定邮箱进行告警，并支持查询历史邮件告警日志记录，便于管理员第一时间掌握同步故障。1.2**非结构化数据采集**针对图片、pdf、office文件、多媒体文件等文档类对象进行自动化批量采集，存储到分布式文件系统中，并进行元数据提取和分类编目，将文档对象与数据仓库中的结构化数据进行关联，提供按类目检索和关键字检索等功能。用于实现非结构化文档对象的持久化存储、资产化管理和对外开放服务，帮助应用程序获取文档对象支撑其业务运行。要求具备访问鉴权功能，确保获取的文档下载链接仅在指定授权时段有效。**1.3线下数据采集**支持电子表格数据进行采集入库、标准化处理、线上化维护的能力，使非专业数据操作者能够对这些数据进行便利的查询、维护、更新，并使这些数据成为可在全校共享流通的线上数据资源。支持对指定的字段进行格式或规范约束。当用户在该字段中填入或导入不符合规范的数据时，系统应拒绝该数据的提交，并给出明确的提示。为保障用户填报或导入数据符合数据标准的代码定义，产品需支持对指定的字段绑定数据标准中定义的代码表。绑定后，当用户在填入数据时，系统应提供下拉选项供用户选择输入。批量导入数据时，若数值与代码不符，系统应拒绝该数据导入。支持多个部门使用同一个表格进行数据维护，且需支持对不同部门的数据操作权限进行必要的隔离。即在同一表格下，某部门下的多个用户可以共同维护本部门数据，但不得查看和操作其他部门的数据。在同一表格中，若部分数据的管理权发生变更，需支持管理员手工调整这些数据的部门归属，使数据的操作权限从原部门转移到新部门。支持对每一行数据自动化添加操作者的个人标记。管理员可以基于该标记确认该行数据是由哪个用户录入或导入。为实现多人协同维护同一张表的不同信息项，系统需支持对于同一张表的多个字段，对不同用户分别授权可查看、可编辑的字段范围。授权后，用户在该表中只能查看具有查看权限的字段，只能查看和编辑具有编辑权限的字段。支持对多值字段进行自动拆分处理，以提升数据的规范性。用户可以继续按照过往的填报习惯将多个数值填入一个字段，系统需支持基于分隔符将一个字段中的多个数值拆分成为多行数据，并在新行中自动继承其他字段的取值。需支持对一张表的多个字段同时进行拆分处理。需支持将拆分后的数据通过与其他表的数据进行匹配关联，将其他表的若干字段引入该表，成为组成表的数据项。当其他表中被关联字段的内容发生更新时，该表中对应字段信息需自动更新。为提醒数据填报者及时进行数据填报或维护，系统需支持对指定的填报者以邮件或企业微信的方式自动推送消息提醒，支持周期性提醒。支持管理员对表格中的全部数据进行查看、编辑、授权、导出。**1.4数据源管理**支持技术架构先进、速度快、稳定性高、可靠性高的数据库连接池，并在投标文件中须提供对应的功能截图；支持添加、编辑和删除关系型数据库（MySQL、Oracle、SQLServer、PostgreSQL、达梦数据库等）、非关系型数据库（Hive、Kafka、MongoDB、Redis等）、Hadoop平台（CDH、MRS、TDBS等）数据源；若添加的数据库属于数据仓库本身自用的存储容器，要求支持一键标注，使该数据库注册为数据仓库；配置数据源所属部门、科室、数据源类型、业务系统、系统厂商等信息时，要求能够与元数据进行联动，支持自动通过下拉框的形式进行相关信息的选择即可完成配置，而无需手工输出，确保元数据的唯一性和准确性；添加数据源时，支持进行最大连接数、最小连接数、链接超时时间、最大回收时间、校验时间、空闲超时时间的按需配置；为提高元数据的采集效率，支持对数据源中的数据进行元数据采集任务的一键配置。为满足国产化和信创要求，数据管理平台软件需适配至少2种国产处理器、国产服务器操作系统，要求提供兼容性认证证书。数据管理平台软件需适配至少三种主流国产数据库。支持按照数据源连接名、数据库类别、数据源类别等对数据源进行筛选查询；支持选择任一数据源查看详情，包括数据源添加的配置详情及该数据源下包含的数据表信息，支持按照表名称、数据标准、对象类型和所属部门进行数据表的筛选查看，选择某张表后支持进一步查看数据表详情和数据详情；支持对各类用于接收数据的中间库进行统一的管理和配置，并可查看详情、测试连接状态；可查看数据申请者创建的中间库的数据源连接名、数据源类型、所属部门、所属科室、已被申请的ETL接口数。支持对数据源执行停/启用操作，数据源停用后将不再提供数据支撑。即其他关联模块相应数据立刻暂停服务。如学校的业务系统停用后，数据资源将不再为数据中心更新数据，投标文件中须提供对应功能的产品截图。支持批量导入、导出数据表的数据字典，支持通过导入数据字典填充数据表的注释、字段的注释。为了便于数据中心管理员进行数据资产的灵活管理，要求提供数据询功能，让管理员能够对各类数据源中的数据详情进行按需检索和查询。支持通过数据源类型、数据库连接名、schema的信息进行数据源的筛选，筛选后支持根据中文注释和表名称对数据表进行查询，检索结果支持导出为Excel或进行全屏查看；选择数据表后支持自动生成SQL查询语句进行数据详情的查看，也需支持用户通过自定义SQL语句进行查询，支持对查询到的数据详情进行更详细的高级检索，支持按照字段名称、筛选条件和值进行检索，支持对非关系型数据库和Hadoop平台的各类数据源进行查看，支持查看的数据源类型包括但不限于Hive、kafka、MogoDB、HDFS等，支持查看的内容如HDFS的文件情况；Kafka数据库的Topic等。**1.5数据集成报告**为便于管理人员快速掌握数据集成的整体运行情况，可通过数据集成报告统计运行情况，报告页面展示内容至少有：A.ETL数据集成的整体运行情况，包括当日运行异常监测、管理端接口运行情况、数据交换量统计、调度运行情况、接口耗时排名排名、接口失败次数最多排名、数据交换趋势、数据源流量。B.线下数据集成的整体运行情况，包括：离线数据采集情况、数据采集量、导入数据量排名排名、手工填报量排名排名、部门数据表填报量统计、填报数据量趋势。 |
| **2** | **数据管理** | **2.1元数据管理**支持查看数据表、视图、存储过程、函数等多种资源对象。支持对表结构的变化自动生成历史版本，支持比对查看差异。要求可对数据对象的来源系统、来源部门等属性进行手工设置，可将一张表整体归属到同一个部门，也支持将一张表的不同字段分属到不同部门，实现字段级的权威数据源登记。可自动构建生成整个数据仓库的全局数据流向图，以呈现整个数据仓库的数据上行、下行流动信息。流向图中，应可呈现从每个部门、每个业务系统采集到数据仓库的数据量、数据内容和接口方式，以及数据仓库向每个部门、每个业务系统提供数据服务的数据量、数据内容和接口方式，流向图中，可以对从各部门、各业务系统采集的数据内容进行下钻显示详情，包括每个采集接口的运行状态、源库/表、目标库/表、最近执行时间、最后一次数据变化量等信息，并可以直接定位到每个ETL接口的管理界面。可以对数据服务的内容进行下钻显示详情，包括每个服务接口的运行状态、数据清单名称、源库/表、目标库/表、最后执行时间、最后一次数据变化量等信息，并可以直接定位到该数据清单或ETL接口的管理界面。可自动构建生成单个数据对象的微观数据链路图，自动分析并显示数据对象的来源部门、来源系统级、来源表、来源字段，下游服务的各项应用，每个应用的名称、所属部门、调用的字段范围。要求在所有节点上提供字段名称检索框以便快速定位到特定字段，需支持模糊匹配。提供数据明细内容的预览和查询入口。可进一步检索每个字段在全链路上各个节点处的具体取值，以便查看指定的字段取值是否存在异常变化或数据丢失。支持对系统内需要使用的厂商名称、业务系统名称、业务部门的信息进行管理，支持新增/编辑/查看/删除/批量导入操作，并根据数据的实际采集来源，将这些信息自动关联到数据资源上，用来标定数据的来源系统、来源部门。支持将数据治理服务过程中产生的所有过程性及结果性文档进行线上化管理，同时支持将配套的数据管理制度进行线上化管理。具体包括：A.数据治理启动前期可以实现业务部门信息化接口人信息表、业务系统数据字段、数据库只读权限等文档的线上化管理，并支持对同一文档的不同版本进行线上化管理，支持文档的上传、在线预览、下载和删除。B.调研阶段支持将调研过程性文档（数据共享交换需求、数据分析需求、数据更新频率需求、线下数据线上化管理需求等）和结果性文档（数据标准、代码标准、编码规则、数据权责等）进行线上化管理，并支持对同一文档的不同版本进行线上化管理，支持文档的上传、在线预览、下载和删除。C.运维阶段支持对数据管理制度、用户操作手册、开发者手册等过程性文档进行线上化管理，支持文档的上传、在线预览、下载和删除。可按模块根据关键字对元数据进行检索。检索结果可按技术元数据/业务元数据/管理元数据分类展示。可配置元数据自动采集任务，支持手动执行采集任务。支持对采集任务进行自定义编辑。支持批量执行/删除元数据采集任务。可查看元数据采集的表名单。发生元数据变更时，可针对指定邮箱进行邮件预警。**2.2数据标准管理**可对标准数据元的集合进行在线管理。支持导入、导出标准数据元/代码集/编码规则，其中标准数据元应在数据模型中强制使用。支持为数据元绑定代码表、质量检测规则、编码规则、脱敏规则。绑定完成后，当某个模型引用了该数据元时，其绑定关系应自动继承到的模型中生效。可生成数据元的链接图谱，图谱中可查看该数据元的各项属性、关联的模型、关联的码表、质量检测规则、编码规则、脱敏规则。支持对数据标准进行全局检索。支持对数据元/代码集/编码规则进行批量上线/下线操作。要求可对不同版本的标准进行比对及手动备份。可监控业务系统的数据模型与数据标准之间的差异度。可监控数据标准被引用情况，需支持自定义添加监控名单。支持代码标准与业务库非标准字段对比检测，并提供图形化界面在线对代码标准进行映射、维护与管理。支持通过数据报告页面快速查看当前数据标准模块的整体运行情况，展示内容应包括：数据标准建设概览，包括公共属性集合、代码表、编码规则分别进行统计；近三十天标准的变化趋势；标准检索热门关键词排名等；**2.3数据模型管理**支持在线查看及编辑标准数据模型，定义数据模型的字段组成、英文名称、中文注释、相关联的实体表、版本数量等信息。支持按照主题对标准数据模型进行分类管理。支持自定义分类，不限分类层级。支持对单个标准数据模型管理的内容变化进行自动跟踪，每当内容发生变化时，自动生成新的版本，并留存历史版本的内容。可随时查看每个版本的时间和具体内容。支持基于标准数据模型直接在数据库中自动生成对应的实体表，也可以生成SQL脚本以便手工建表。要求生成实体表的操作既支持针对单个模型，也支持一次性针对多个模型批量生成实体表。基于数据模型自动生成实体表时，系统应自动记录模型与实体表的关联关系。支持手工断开模型与实体表的关联关系，断开关联关系后，允许利用同一个模型重新生成另外的实体表。支持对模型与实体表各自的版本变化进行动态跟踪。当模型与实体表出现差异时，系统应自动予以提示，并给出具体的差异详情。支持对单个标准数据模型管理的内容变化进行版本跟踪，每当内容发生变化时，自动生成新的版本，并留存历史版本的内容。支持对单个数据模型的不同版本进行内容差异比对，比对后可显示存在差异的字段名称、字段定义等详细信息。支持对标准数据模型的集合进行版本管理，指定主版本作为生效版本。可对标准数据模型集合的不同版本的进行内容差异比对，比对后可显示存在差异的模型名称、字段定义等详细信息，可批量导出差异信息。支持在线管理所有由标准数据模型生成的实体表对象，查看实体表的字段组成、英文名称、中文注释、存储位置等信息。可显示每个实体表与其对应模型的一致性，存在差异时，系统应自动予以提示，并给出具体的差异详情。可在线编辑实体表，进行字段的新增、删除、修改。在数据模型设计过程中，要求系统自动强制保证标准数据模型中仅允许引用在数据标准中已经实现定义过的标准数据元，并自动继承标准数据元的各项属性、各种绑定要素，从而保证所有标准数据模型中的字段定义具备全局一致性。要求系统自动对数据模型与实体表的一致性进行动态跟踪。当存在差异时，系统应自动予以提示，并给出具体的差异详情。支持一键操作完成模型与实体表的同步，既支持修改模型以匹配实体表，也支持修改实体表以匹配模型。在实体表编辑过程中，支持一键操作将实体表的字段变更同步应用到对应的模型上，从而维持模型与实体表的一致性。同步过程中，实体表中新增的字段可自动扩充到标准数据元集合中，从保持标准数据元素、标准数据模型、实体表三者的一致性。**2.4数据工程看板**自动收集数据，呈现数据治理成效，建立多维度的评估体系，从经验视角到数据视角的数据管理，有效管控数据治理成效。记录数据的流向，可实现数据的溯源，如：业务部门数量、业务系统数量、厂商数量、数据表等，实现数据治理成果溯源管理。实现治理成果的监控管理，看板指标每日自动刷新，也支持手动即时刷新。可导出治理成果的监控指标。导出时可选择具体日期，导出后可查看业务系统、数据湖、数据仓库、数据集市、数据共享情况，以表格形式对各项值进行统计。可查看近一个月、近三个月和近一年的数据统计变化趋势，支持折线图的展示，便于查看管理，对治理成效进行管理。通过报表统计记录库表对接统计、数据补录统计、各部门数据交换情况，可根据时间检索，可导出统计数据。**2.5数据仓库运维**可查看数据源管理模块的数据概览情况，包括：异常数据源、停用数据源、删除数据源、DB连接用户变更数，支持查看详情。可查看数据集成模块的数据概览情况，包括：ETL接口异常数、删除数据表、任务异常数和敏感操作数，支持查看详情。可查看数据目录模块的数据概览情况，包括：主题清单异常数、今日申请数、已处理申请数，支持查看详情。可查看数据标准模块的数据概览情况，包括：数据元变更数、代码表变更数、编码规则变更数，支持查看详情。可查看数据模型模块的数据概览情况，支持查看实体表异常数、删除实体表数、敏感操作数，支持查看详情。可查看元数据中心模块的数据概览情况，包括：元数据结构变更数、元数据删除数、元数据采集任务失败数，支持查看详情。可查看主数据中心模块的数据概览情况，包括：主数据异动数、主数据备份失败数、敏感操作数，支持查看详情。可查看数据质量管理模块的数据概览情况，包括：检测规则删除数、检核任务失败数，支持查看详情。可查看数据安全模块的数据概览情况，包括：敏感操作数，支持查看详情。可查看服务器资源总览，包括：磁盘利用率、CPU利用率、内存利用率。可查看数据中台中业务系统的基本信息，包括：业务部门数、业务系统数、厂商数量、数据表数量、包含数据项、数据总量、数据总体积、表注释率、字段注释率；可查看数据中台中数据仓库的数据治理成果信息，包括：标准模型数、业务部门数、业务系统数、数据表数量、包含数据项、数据条目数、数据总体积、代码表数量、代码表数据项、代码表数据条目数、检测数据表、检测数据项、已绑定检测规则数、检测规则数、检测规则利用率、上行接口数、下行接口数、开放数据清单、数据申请次数、服务应用数、服务用户数、空值率、表注释率、字段注释率。 |
| **3** | **数据共享** | **3.1数据资产目录管理**提供数据资源目录展示所有已发布的数据资源。至少显示数据资源的所属部门、所属业务系统、数据接口格式、字段说明、申请总数、更新频率、字段数、调用次数、数据量等信息。支持进行数据预览功能，预览时自动对数据进行脱敏显示。**3.2数据资源发布管理**1）支持将多种不同类型的数据资源发布到数据资源门户中，供用户按需申请所需数据。支持自定义资源的整体分类方式和类型名称，支持自定义多个分类层级。发布数据资源时，支持自定义资源名称和所在分类，支持资源配置检索关键字便于使用者进行查询。2）支持对结构化数据资源进行发布，至少支持API接口、离线表格、数据库、ETL4种发布方式。并支持对发布的数据清单进行脱敏、加密等操作。包含附件字段的，支持指定附件字段是否允许下载功能。支持选择数据资源对外服务时暴露的字段范围，支持选择全部或指定部分字段。对于API接口方式，支持选择API的封装格式，且至少支持JSON、XML、TXT三种格式。支持对数据发布的范围进行再次编辑，至少应包括增加字段、注释更改等操作。在结构化数据资源已发布为API接口并被使用后，正常运行的情况下，支持编辑过程中API接口服务不中断，下游应用无感知。3）支持对非结构化的文件资源进行发布。至少支持发布文本、OFFICE文档、PDF文档、图片、音视频媒体、压缩包、XML等不同类型的资源，支持的文件格式类型不少于25种。4）支持对第三方API接口进行发布，实现API的转发服务。支持指定API请求参数，至少包括定义参数的名称、参数位置、数据类型、默认值等规范。5）支持对第三方KAFKA消息队列进行发布，实现KAFKA消息订阅服务，发布时支持指定特定Topic。6）对已发布的清单，提供修改清单状态、进行数据测试或删除数据清单等功能。对已发布清单的所属部门、主题提供批量设置功能。支持下载数据清单的数据字典；提供数据清单导出为Excel功能。支持多种排序规则对数据清单进行排序，排序规则至少包括api调用次数、离线表格下载次数、支撑应用数、申请数等。**3.3数据资产服务的安全策略管控**总体要求：具备策略化的数据资源服务的管控能力，实现精细控制数据资源对应用提供服务的细节，至少支持从调用主体、调用频次、数据范围、服务时间等多个不同维度对数据资源的开放服务进行有效管控。1）支持使用https协议承载API数据服务。2）支持数据加密和脱敏功能。支持发布数据清单时，可以加密或脱敏选项保障数据服务的隐私安全。在数据门户界面下浏览数据清单时，支持在获得正式授权之前，只能查看经过脱敏的数据内容功能。3）至少支持AES、MD5、国密算法（SM2-4）加密方式。脱敏方式至少支持数字随机脱敏、字符仿真脱敏、字符随机脱敏三种方式，提供在数据发布审核阶段按照数据字段颗粒度进行加密和脱敏功能。4）发起数据资源申请时，提供先查看数据使用保密协议，确认后才可以进行申请功能。支持针对特定应用控制其调用某个数据资源的内容范围，可设定判断条件进行筛选控制，支持组合多个判断条件进行复合筛选，或通过SQL语法进行自定义筛选。支持手工取消某个应用对数据资源的调用授权。5）支持对数据资源服务的调用频次设置上限，可按照小时、天、周、月等不同时间颗粒度分别进行设置。支持设置数据资源的调用频次限制时，既可以作为全局性限制对所有应用生效，也可以作为针对性限制仅针对特定应用生效。设置全局性调用频次限制时，至少可以根据分类、关键字等属性一次性选择多个数据资源，设置后应自动对选择范围内的所有数据资源生效。设置调用频次限制后，提供某应用程序实际调用数据资源的频次超出策略设定的上限时，应自动停止响应该应用对数据资源的调用功能。管理员可查看超出了频次限制的具体应用程序和数据资源组合，并允许进行手工重置以便恢复数据服务。支持设置数据资源的有效访问时段功能。至少支持指定允许访问时段和指定禁止访问时段2种规则模式，支持按照天、小时、分、秒等不同时间颗粒度设置数据资源的允许访问时段或禁止访问时段。支持设置有效访问时段时，既可以作为全局性设置对所有应用生效，也可以作为针对性设置仅对特定应用生效。设置全局性有效访问时段时，可以根据分类、关键字等属性一次性选择多个数据资源，设置后应自动对选择范围内的所有数据资源生效。设置数据资源的有效访问时段后，应用只允许在有效时段内访问数据资源。支持配置多个不同内容的有效访问时段控制策略，同时生效执行。1. 支持数据访问的黑白名单控制。支持白名单（指定允许名单）、黑名单（指定限制名单）2种规则模式。支持在白/黑名单中指定单个IP地址或IP段范围或域名。可以根据分类、关键字等属性一次性选择多个数据资源，对选择范围内的所有数据资源同时生效。支持配置多个不同内容的源地址控制策略，同时生效执行。

上述各种控制操作，如果是针对特定场景生效的须对每个场景及其相关的数据资源独立生效，互不影响。1. 使用数据资源时，提供线上审核功能，通过流程审核后才能被调用。系统预定义多种用户角色用于数据资源的申请和审核权限管理。系统预定义多种审核流程模式。每种流程模式中，至少支持指定流程节点的数量、每个节点的审核主体、流程的执行顺序等特征。
2. 支持对数据设置安全等级，颗粒度可达到字段级别。支持基于安全等级，实现不同安全等级下数据审核流程的多种不同模式。
3. 具备常态化的数据资源服务状态的监控能力，实时展示各类数据被使用的现状及趋势。

支持监控每个应用当前已经审核获批允许访问的数据资源的详情，至少包括该应用使用了哪些数据接口类型，每种类型下分别包含哪些资源对象，每个资源的申请审批记录，每个审核节点的详情信息，每个资源下授权了哪些字段，各个字段是否使用了加密、脱敏的安全控制策略，是否允许回写数据等等。支持对API接口可用性进行测试，验证接口的可访问性和数据可用性。支持监控每个应用实际调用数据资源的详情，至少包括该应用对每个资源调用的次数，可查看调用频次的趋势图。对数据资源提供服务时的性能进行监控，提供检查每个API数据接口在数据库查询阶段和API封装阶段的时间消耗功能。支持自动扫描检查每个数据资源的服务有效性。支持自动化定期扫描或手工执行扫描。扫描后可查看扫描执行的时间、扫描消耗的时长、扫描的数据资源总数以及其中正常/异常的数量，并可以查看异常资源的具体名称和发生异常的原因。当出现数据资源服务故障时，支持监控每个数据资源提供服务时出错的详情。至少包括数据资源的名称、应用的名称、发生故障的时间点、错误原因和错误日志的明细内容。支持对数据清单接口调用失败进行监测，提供查看接口调用失败的记录能。 |
| **4** | **数据门户** | **4.1运行状态监控：**至少支持接口超限、访问、调用异常等维度监控，支持查看应用调用数据超限详情，至少包括每小时或每天、每月超出限额的情况。**4.2资产目录总览：**支持查看数据资产目录的总览，至少包括数据来源部门及系统，发布数据清单、代码表、数据项等个数，支持查看数据量、累计申请次数、为几个业务部门及应用系统提供了数据支撑等。支持查看数据集市建设情况总览，至少支持查看数据集发布、下载、申请情况的热度排名。**4.3数据资源服务情况展示：**展示校级数据资产服务情况须提供情况展示功能，至少包括全校数据资产概览、数据共享交换情况、数据供需情况、数据质量情况等；包括但不限于采集了多少个业务部门多少个生产库的数据、已完成了多少个数据标准模型、代码表、编码规则的建设、有多少个符合标准的数据清单能够对外开放，且各个模块均支持下钻展示更详细的信息；支持对业务部门进行下钻展示详细的部门名称、各部门需要遵循的数据标准表有多少个、各部门根据数据标准需提供的数据清单个数、已提供的个数及未提供的个数；支持对生产库进行下钻展示详细的系统名称、各系统归属的管理部门，各系统的原始表数量及已提供的表数量；支持对数据标准模型、代码表、编码规则进行下钻展示数据标准模型数量、标准模型涉及的数据项数量及涉及的部门数量，展示代码表的数量及涉及的字段数量，展示编码规则数量及被调用情况；支持对数据资源进行下钻展示资源提供情况（包括已/未提供表及涉及的字段数）、资源的重要性（包含重要资源表及重要信息项）**4.4数据共享交换情况展示：**1）支持在一个图中展示学校所有业务部门与数据中心的数据共享交换情况，能够直观展示每个业务部门共享至数据中心的表数量、支持下钻查看详情（包括下钻后的数据检索功能、数据在部门端的表名称、表所属业务系统，及共享至数据中心后的表名称）；2）支持展示数据中心通过不同方式（ETL方式、API方式等）交换至每个业务部门的表数量、支持下钻查看详情（包括下钻后的数据检索功能、数据在数据中心的表名称、交换后的目标业务系统、表名称及数据最近更新时间）；3）支持汇总展示数据表总量、已共享交换表数量及涉及交换部门数量；4）支持对业务部门进行下钻展示部门负责的数据表个数、字段个数，数据异常趋势图、异常数据百分比及已提供数据百分比；5）支持针对单个数据对象查看该表的全链路分析，提供该表从来源采集、生产加工、资源发布及被应用调用后的全局数据流向图，支持查看表级、字段级数据流向展示功能。**4.5部门数据展示：**支持汇总展示全校各部门已/未提供数据字段数量，支持按部门展示已/未提供数据字段数量及提供率；支持直观展示部门模型确认情况排名：包括归属某部门职责范围内需要确认的标准模型总数量、确认无误数量、确认异常数量及尚未确认数量；**4.6数据质量情况展示：**支持汇总展示当日数据质量检测情况，至少包括正常数据、异常数据及合格率等；支持汇总展示全校数据治理概览，至少包括总体概况（表数量、信息项数量、数据量）、质量情况（问题表、问题信息项、问题数量）、处理情况（已解决数量、未解决数量、新增问题数量）、检测情况（检测规则数、检测表数、检测字段数）及重要字段（重要字段数及重要字段问题率）。支持直观展示各部门数据治理排行，展示内容至少包括部门名称、总检测项、异常项、异常率； **4.7数据标准查看：**支持通过数据服务门户查看数据标准模型、代码集、编码规则等相关信息。支持数据标准模型按照模型分类或检索查看某张具体模型表的详细信息，至少包括模型表名称、归属的责任部门、最近更新时间、字段数量及字段说明信息；支持代码表按照分类或检索查看某张代码表的详细信息，至少包括代码表名称、代码记录数、被引用次数、最新更新时间、发布时间，支持下钻查看代码详细及代码的下载，支持新建代码集；支持编码规则按照分类或检索产看某个编码规则文档详情，至少包括编码规则文档名、被引用次数、最新更新时间及背部时间，支持下钻查看编码规则文档详情、文档下载及提交编码规则使用申请；**4.8数据集市：**1）支持前台用户对数据清单进行检索，支持查看集市清单的详情，至少包括：字段说明、数据预览等；支持对API接口数据进行测试，验证API接口返回内容的有效性、并支持转换为json2数据进行查看，支持查看接口参数约定和调用API示例。在查询的数据量为1万条以上记录的情况下，1万并发用户查看单条数据的响应时间小于1秒。数据共享开放的数据量为50万时，请求响应时间小于2秒，数据交换耗时小于60秒。**4.9申请数据资源：** 数据资源目录中支持自选数据资源，提供在线提交使用申请功能。申请时可指定所需的字段范围，支持在线进行数据使用保密协议的签订。提供购物车模式，支持批量申请。支持对数据资源进行收藏，对API接口提供数据测试验证连通性和有效性功能。**4.10审批数据申请：**数据用户发起数据资源申请后，系统支持预定义的审核流程模式自动生成相应的审核流程，并在相关审核人的界面中予以提示，提醒尽快进行审核处理。支持在审核流程流转过程中，系统支持申请的各个字段的不同来源部门，自动将各个字段的审核权限呈现在其对应来源部门的审核界面中，无需一一人工指定。一次审核中涉及多个来源部门时，系统支持自动将流程平台投送到多个部门。在审核流程流转过程中，某个审核节点完成审核动作后，系统支持预先定义的审核流程模式，自动将审核事项流转到下一个审核节点，直至整个审核流程执行完成。审核时，支持审核者对申请者提交的字段范围进行筛选，支持设置数据的筛选条件，支持多条件逻辑组合；支持设置数据资源的有效访问时段，时间段的颗粒度至少精细到一天；支持对特定字段启用加密和脱敏策略。支持只有审核流程全部通过并结束后，申请者才能获得数据资源的连接参数，使用数据时，系统至少支持审核时获批的字段范围、筛选条件、访问时段、加密脱敏、有效源地址等限制策略，确保数据申请者在上述限制条件的控制下连接并访问数据资源。**4.11部门目录：**至少支持按部门查看各部门的数据清单、采集清单、标准模型。支持按照部门视角查看数据流向地图，流向包括但不限于数据从业务系统到ODS、DW及数据被应用调用的情况，支持下钻查看ETL、API等类型接口的详细信息，ETL接口详细信息包括接口状态、接口名称、接口来源、最近更新时间、源库及表名称、目标库及表名称； 支持按权限查看标准模型的确认状态。支持按权限查看本部门与数据中心的数据交换情况，及交换涉及的集成接口、集市清单；支持按权限按照不同维度（包括按系统、重要性、上架状态、开放状态、安全级别）查看本部门数据资源情况，支持按权限导出数据资源目录及下载部门数据字典，支持下钻查看某张表的字段说明、数据预览、数据溯源及数据互动（包括建议或意见、数据纠错、质量需求及其他需求反馈）；支持按权限查看本部门的采集清单，支持采集清单下载到本地；支持按权限查看本部门用户申请的数据情况及对其他部门提供的数据情况；支持按权限按部门查看各部门的数据清单、采集清单、标准模型。**4.12对已申请资源的管理：**支持查看本人发起的的申请记录，至少包括待审核、审批通过及审批拒绝的记录详情；支持查看本人申请获批的各类数据资源的详情，至少包括API接口类、文本类、ETL接口类、DB类数据资源。 支持查看已创建的应用详情，至少包括应用名称、申请的接口数据、文本数据、ETL数据、DB数据、离线数据、数据标准及应用状态，支持查看、编辑、上架、删除及查看应用授权码；支持查看API接口的调用类型说明、授权方式、调用频率设置、错误码说明信息查看及调用示例信息查看等。支持查看所有已提出的数据需求记录被审核情况，包括待审批、审批通过及审批拒绝的记录详情等；提供数据详情查询界面，支持对数据资源的内容进行在线查看、检索。支持基于中文语义的检索条件设置，至少包括等于/不等于/有值/无值/开始于文字/结束于文字/不包含文字等各种条件。提供在线反馈功能，可查看新的数据需求、数据质量问题等。平台支持不低于500用户并发查询，平均响应时间小于0.3秒，事务成功率100%；支持不低于1000用户并发查询，平均响应时间小于0.5秒，事务成功率99.9%。提供第三方机构出具的检测报告复印件或官网截图等有效证明材料并加盖公章。**4.13 API托管功能：**提供第三方API托管功能，支持图形化界面交互式操作。托管代理过程中，支持API请求者发起API调用时，采用数据中心统一规定的鉴权方式；数据中心向API供应者发起API调用时，采用API供应者要求的特定鉴权方式，二者相互隔离互不影响。托管代理过程中，要求API的数据结构不发生变化。即API请求者向数据中心调用API，与直接向API供应者调用API时，返回信息的结构、内容完全一致。整个过程须自动化、透明化完成，实现近实时的API代理服务。 |
| **5** | **数据标准制定服务** | 校级数据标准1要求根据国家标准、代码标准，教育部标准、代码集，以及行业标准，并结合学校业务和数据特点进行归纳整理，建设一套符合学校实际情况的《元数据标准模型》、《标准代码集》；数据标准建设条目不少于 100 个；2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；制定有数据标准、数据交换标准、数据集成标准和接口集成规范等管理文件；能实现数据标准管理，有明确的数据标准管理办法和应用流程；3标准的制定需根据我校已形成的事实数据标准，本着尽量减少对学校业务系统，管理过程的影响为原则，尽量符合国家标准、教育部标准为前提条件形成我校数据标准初稿。4依据我校数据治理范围构建元数据子集相应的代码标准，每个元数据子集都对应着相应的标准代码，以及代码的定义与说明。5依据我校数据治理范围制定学校自定义编码规范，提供相关规范文档。 |
| **6** | **数据交换迁移服务** | 旧数据共享交换体系迁移1.基础数据采集迁移至新建平台。涉及人事，教务，一卡通，图书馆等基础数据。2.新建平台与各对接应用系统数据共享交换。主要为基础数据。 |
| **7** | **统一身份认证对接** | 数据门户与统一身份认证对接，实现单点登录。 |

一期项目招标硬件技术要求

|  |  |  |
| --- | --- | --- |
| **序号** | **配置** | **详细参数** |
| **1** | **服务器数量** | 2U机架式服务器2台，带导轨。 |
| **2** | **CPU** | ≥2颗；频率大于2.3GHz;性能不低于Intel Xeon Gold 5218 16核； |
| **3** | **内存** | ≥2\*64G DDR4 3200；32个内存插槽； |
| **4** | **硬盘** |  4块1.2T SAS 10K 机械硬盘， RAID5保护数据安全。 |
| **5** | **网卡** | 4口千兆以太网卡,支持远程WEB管理功能。 |
| **6** | **电源** | 2个550W电源。 |