

延安职业技术学院校级数据服务平台和大数据分析（一期）建设项目(二次)中标 (成交) 明细

受延安职业技术学院委托，采用进行采购校级数据服务平台和大数据分析（一期）建设项目(二次)（项目编号：YAZCJC2022-32-1）项目，中标（成交）供应商名称及中标（成交）结果如下：

一、合同包1（延安职业技术学院校级数据服务平台和大数据分析（一期）建设）

1.1、中标（成交）供应商：南京迪塔维数据技术有限公司

1.2、中标（成交）总价：1069000.00 元

1.3、中标（成交）标的明细：

服务类

序号	品目名称	服务名称	服务要求	服务时间	服务标准	数量	计量单位	单价(元)	总价(元)
			<p>应用系统平台开发要求：（1）服务中心功能主要实现对服务中心进行高可用高性能的集群化管理，实现通过服务配置管理功能，对服务中心进行集中和动态的方式管理所有服务的配置信息。实现通过服务注册管理功能将项目上线涉及发布的服务进行统一管控，填写服务的基本信息并与项目信息进行关联，填写服务相关的运维人员信息、接口文档、部署文档等。实现通过服务发现管理功能完成订阅服务的管理以及服务节点的权重设置、服务上下线处理。实现服务注册、服务配置、服务发现等管理功能，实现服务集群节点信息的添加、删除、下线、上线功能，其中集群节点信息包含服务中心节点IP，服务中心节点状态。（2）网关服务管理主要实现对后端接口服务进行多个目标节点进行负载均衡和健康检查，其中网关服务管理功能包含对后端接口服务进行配置多目标节点，设置服务的负载均衡算法，以及配置健康检查规则，并对已配置好的网关服务进行多条件组成查询和维护操作。平台具备网关服务查询、网关服务新增、网关服务修改以及网关服务删除等功能。（3）网关路由管理主要实现作为所有网关服务请求的统一入口，实现通过网关路由自定义客户端请求与服务之间的匹配规则，以及支持与微服务、负载均衡服务关联，实现一个网关服务可以对应一组路由，一个路由可以对应一个网关服务，每一个匹配到路由的请求都将被网关代理到路由绑定的网关服务中。（4）网关鉴权管理主要实现应用接入信息的管理，应用授权访问可调用的网关服务范围配置、IP黑/白名单的配置、以及网关服务鉴权功能。（5）提供应用网关负载均衡算法库，主要实现对各种常规负载均衡算法进行定义，并对网关服务、网关路由提供负载均衡能力的支撑；提供网关组件库，其主要实现对常规的网关组件进行定义和组件参数的维护，并对网关服务、网关路由提供鉴权能力，限流控制能力，服务缓存能力，日志收集能力。（6）平台需实现能够对用户请求发起到结束的整个过程中，这个请求经过的整个链路上的所有被访问到的服务打上一个相同的标记，能够通过该标记，可以追溯到链路上下游所有</p>		<p>1数据运维模块。数据运维模块作为数据中台的运维数据汇总，具备整体数据运维情况呈现，统一身份认证对接、应用管理功能，并具备数据中台指标管理、标签管理等核心功能。2数据治理模块。数据治理模块集成各类业务系统数据，对数据进行清洗比对加工、以及建设统一的数据标准。可面向应用提供高价值数据支撑，确保大数据应用分析的准确性。数据治理模块包含数据地图、数据查询、数据血缘、数据管理、数据集成、数据标准、元数据管理、数据一致性检查、数据比对、数据脱敏、质量稽查对象、质量调度规则管理、质量稽查规则管理等功能。3数据共享交换模块。数据共享交换模块作为各系统数据交换的桥梁，提供不同数据库、文件</p>				

序号	品目名称	服务名称	服务范围	服务时间	服务标准	数量	计量单位	单价(元)	总价(元)
1	支撑软件开发服务	数据中台软件、数据可视化应用开发工具、业务系统对接、数据治理服务	<p>的调用,至少具备业务理点、线上异常排查、上下游服务依赖分析已经接口耗时分析等链路追踪能力。数据中台基础平台开发要求:</p> <p>(1)数据中台数据层基于hadoop生态构建,数据治理层基于java主流框架springBoot,dubbo微服务框架,前端基于react框架开发。基础平台提供了大数据相关的组件,包括hdfs, yam, hive, zookeeper, hbase等。基础平台通过master-slave架构,实现了在一台机器上管理集群中所有机器,以及机器上的各个组件。对于需要动态调整服务,在web端执行相应的操作,把对应的指令发送给agent。Agent在本地部署新的服务。如果需要添加新服务器,只需要在服务器上安装agent服务,再添加到平台的集群管理中。新机器添加完成后可以在这台服务器上添加新的服务。(2)数据采集能力要求:结构化数据,例如mysql、oracle、sqlServer等数据库使用dataX和Sqoop进行数据采集、数据可存储到mysql、oracle、hive数仓。非结构化数据,例如日志数据,基于flume进行采集,数据可输出到kafka中进行,数据存储到hive。非结构化数据,例如图片、视频、音频、文档;通过hdfs文件管理界面,进行文件上传到hafs中,再将hdfs文件中的数据进行标签化。标签化结构化数据后存储到hive。(3)平台使用hdfs和elasticsearch进行存储。根据存储的数据类型不同,分开存储。文件类型的数据,存储到hdfs下,字符串类型的数据,存储到elasticsearch下。文件类型的数据,主要包括互联网采集的图片,文件, hive数据文件。字符串类型的数据,主要包括日志, json, xml和html。分布式表格系统以及分布式数据库是基于Hive来实现。Apache Hive™ 数据仓库软件为分布式存储的大数据集上的读、写、管理提供很大方便,同时还可以用SQL语法在大数据集上查询。平台针对数据来源和数据作用,分为三层数据。原始库、标准库和主题库。原始库:结构化采集来的数据。标准库:对原始库中的数据做清洗,获得的干净数据。在数据清洗环节,把不需要的数据做过滤,把不正确的数据做替换,把枚举类数据做标准化。主题库:根据应用需要,可以把多个数据做聚合,获得一个可以直接使用的数据。Hive提供了统一的JDBC的使用方式,供服务调用,使用hsq查询hive下的数据,同时hive也提供了Hive Metastore Server,该服务提供hive下管理的表的schema,为其它计算服务提供支持,可以为对接impala和spark这些分布式计算框架。(4)共享数据中心为所有的数据需求方提供统一的API接口,管理员可以为申请用户设置访问权限和数据域权限,数据使用者在取得API授权的前提下,开发应用客户端调用数据接口获取数据服务的模式,系统将记录接口调用记录,形成接口调用分析表。该模式适合于数据量不太大且实时性要求较高的情况,采用Dubbo组件实现。▲数据中台数据开发要求:(1)数据处理在系统中被分成两个步骤,一是数据过滤,二是数据替换。数据过滤主要是把原始库中的脏数据过滤掉,保证基础库中数据都是满足业务要求的。数据替换的主要功能是把各个业务系统中对同一个属性的描述替换成同一个数据,保证属性描述的一致性。(2)离线开发是数据中台平台提供的一站式大数据开发环境,提供了数据同步、数据开发、发布管理、运维监控的全链路解决方案,可用于构建PB级别的数据仓库,实现超大规模数据集成,通过对数据价值的深度挖掘,实现数据的资产化。业务流程根据业务种类将一批作业组织在一起,并在数据开发过程中以业务为单元开发代码。业务流程内的作业节点可以相互依赖,不同业务流程之间的作业节点也能跨流程依赖。作业(Job)是平台的最小运行单元,支持Shell、Hive、Spark、SparkSQL、DDL等多种作业类型。(3)实时开发要求:数据中台平台应提供的低延迟、高吞吐、高可靠的分布式流数据实时分析工具聚焦于本身的业务逻辑,通过可视化图形拖拽的方式便能快速构建实时计算作业,以及可视化页面向导方式创建实时ETL作业。作业是平台的最小运行单元,支持pyspark、Flink、FlinkSql、python等种实时计算作业类型。实时计算支持多种窗口函数,提供全SQL、自定义Jar程序等开发模式,支持作业运行控制、调试等功能;支持函数开发,便于用户自由的开发实现复杂业务,应对复杂的业务场景,弥补SQL开发的边界;提供可视化拖拽的方式快速配置作业,实现异构数据源之间数据的实时同步,并支持数据预处理。▲数据治理能力开发要求:(1)平台需实现元数据管理功能,提供各类元数据管理,包括:业务元数据、技术元数据和管理元数据,支持元数据的基本信息、属性、依赖关系、组合关系的增删改查操作,同时提供丰富的元数据分析能力,包括:血缘分析、影响分析、全链分析、关联度分析、属性差异分析。(2)平台需提供了一套完整的数据标准管理流程及办法,通过一系列的活动,统一的数据标准制定和发布,结合制度约束、系统控制等手段,实现学校数据中台数据的完整性、有效性、一致性、规范性、开放性和共享性管理,为后续数据质量检查、数据安全等提供标准依据。(3)平台需实现数据质量规则的定义和管理,数据质量规则定义数据质量审核的业务逻辑,并能反馈数据治理问题结果给业务部门进行数据问题处理。平台需预设提供规范检查、逻辑检查、及时检查、重复性检查、波动检查、平衡检查等多种规则。规则管理支持用户自定义分组,并可对定义好的规则进行多角度拓展现,使用户对所建立的规则一目了然,协助用户建立完整的质量规则体系。平台提供通过质量监控,产生和保存质量结果,包括对象名称、发生时间、违反规则、级别、状态等。系统需实现自动生成每个质检方案的明细结果表,并允许用户根据分析需要对明细结果表字段进行自定义,为用户进行丰富多样的数据质量分析提供数据。(4)台需提供完整的资产视图,管理者在平台上可概览企业资产,通过合理的方式管理内部数据和提供对外服务。包括对数据资产的编目,数据资产的检索,以及数据资产</p>	合同签订之日起3个月内完成	之间的数据交换与同服务标准在不同网络、不同操作系统环境中的各类应用系统、数据库之间数据共享需求。数据共享交换模块提供对采集数据的初步整合处理,按照统一的数据标准对采集数据进行清洗、转换,以全量或增量方式加载到数据平台。对数据共享交换过程实行全程监控管理,使数据共享交换始终处于有序管理状态,避免乱交换、重复交换等问题,保证数据的安全性。4数据开发能力及任务调度系统。任务调度系统是一个分布式易扩展的可视化DAG工作流任务调度系统。提供了一个可视化操作任务、工作流和全生命周期数据处理过程的解决方案。任务调度系统解决复杂的大数据任务依赖关系,并为应用程序提供数据和各种 OPS 编排中的关系。解决数据研发ETL依赖错综复杂,无法监控任务健康状态的问题。任务调度系统以DAG流式方式组装任务,可以及时监控任务的执行状态,支持重试、指定节点恢复失败、暂停、恢复、终止任务等操作。5数据可视化应用开发工具。支	1.00	项	1,069,000.00	1,069,000.00

序号	品目名称	服务名称	服务范围	服务时间	服务标准	数量	计量单位	单价(元)	总价(元)
			<p>的应用和分析。平台数据安全要求：（1）服务接入管理通过kerberos来管理。Kerberos简单来说就是一个用于安全认证第三方协议，它采用了传统的共享密钥认证方式，实现了在网络环境不一定保证安全的环境下，client和server之间的通信，适用于client/server模型。默认Hadoop各个组件间无任何认证，因此可以恶意伪装某一组件（比如NameNode）接入到集群中搞破坏。而通过kerberos，可以将密钥事先放到可靠的节点上并只允许有限的访问，该节点的服务启动时读取密钥，并与kerberos交互以做认证，从而接入到hadoop集群中。（2）在用户管理上，通过kerberos, ldap, sentry。管理用户访问服务的权限，用户访问数据的权限。敏感字段，提供了数据加密服务。实现数据的实时动态加密。日志和安全审计，在平台访问中增加了用户访问记录，记录下每个用户访问的页面。并在日志审计页面，展示出来每个用户访问记录和操作记录。详尽的审计日志输出：日志和安全审计，通过Django的拦截功能，可以把用户所有的请求记录都进行记录。在在日志审计页面，展示出来每个用户访问记录和操作记录。提供组件细粒度访问权限控制：平台访问权限管理采用RBAC模型。可以根据角色不用，每个角色定义不同的访问权限。（3）数据脱敏处理：数据脱敏是数据加密的一种表现形式，在数据展示时，既可以选择不同的加密算法对数据进行加密，也可以选择使用“***”对数据进行脱敏展示。（4）敏感性数据加密：数据加密服务设置中，可以设置字段加密类型（sha, rsa, aes）等。在查询数据，数据加密服务根据前台的加密设置，对查询到的数据进行加密处理。并把加密后的数据返回。前台展示的数据是已经加密过的数据。▲平台性能要求：（1）系统具有很好的稳定性、高并发性，平台不限制用户数，软件架构承载支持不少于20000用户并发使用，支持集群模式部署,弹性动态地支持后期增加的并发要求。各微服务应用节点支持横向扩展，采用松耦合构件方式，提供分布式数据库架构设计，可以使数据库的性能可以随着节点增加线性地增加，快速响应业务需求的变化，各逻辑模块可单独升级，不会影响其他逻辑模块的正常运行。（2）平台底层选用集群，支持数据库分布式部署，保证支持10万以上用户注册量。用户访问采用多台服务器负载均衡，Nginx代理访问机制，使前端访问能随机分配到后台多台主机服务中的一台。确保支持多人并发操作，平台能正常运行。（3）对于数据量比较大，业务逻辑复杂，操作繁琐等功能，根据具体应用功能模块，选取对应技术实现查询响应时间小于1秒。如：AP数据量达到TB级别，选取Hive、Solr、ElasticSearch等组件实现查询毫秒响应。（4）对于业务逻辑复杂，计算次数多等情况，使用Hive、Persto、Impala等组件，执行Sql语句操作。后台运行批处理实现数据降维，使Web端操作响应时间不会超过30秒。</p>		<p>持用户自由创建可视化项目对外「公开发布」或「加密分享」，在网络环境互通的情况下，目标用户无需登录系统即可直接查看分享出的可视化项目。6业务系统对接技术要求。对接现有人事、教务、学工、科研、资产、一卡通、线上教学等8个业务系统的数据到全域数据中台软件，并将清洗治理后的标准数据通过API等方式同步给各业务系统，给业务系统厂商带来一定的额外开发工作量，需支付一定的费用。7数据治理服务。以数据共享需求和应用需求为导向，基于全域数据中台软件产品提供标准化的数据服务，以标准、全面和实用为原则，建设学校的全域数据中心数据仓库，形成数据仓库的分层存储，制定符合学校实际及发展需要的数据管理制度及相关数据标准，完成各个业务系统之间的数据清洗、入库，整体盘点数据资产，实现数据一数一源的落地交付。</p>				