### 江南水务数据中台项目询价采购需求

1. **项目背景**

为了进一步贯彻落实江苏江南水务股份有限公司（简称：江南水务）对外服务承诺，提高营业管理及服务效率，开展信息系统新技术研究和应用测试、对外技术合作、引进和评估，吸收契合公司发展需要的新信息化技术，支撑公司不断变化的业务需求。

1. **项目建设目标**

根据江南水务信息化战略，构建江南水务数据中台，实现业务高效协同，信息充分共享，发挥大数据在企业精细化管理和决策中的支撑作用。依托大数据技术实现公司内部数据资源的集中整合，引入外部有价值的各类数据，形成公司统一的数据湖。

1. **采购清单**

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 名称 | 技术参数 | 数量 |
| 1 | 数据中台软件产品 | 详细见下文 | 1套 |
| 2 | 数据中台实施服务 |  |  |
| 3 | 物理服务器 | 详细见下文 | 3台 |
| 4 | 万兆交换机 | 详细见下文 | 1台 |
| 5 | 千兆交换机 | 详细见下文 | 1台 |

备注：

1、签订合同后一周内,供应商须提供软件产品功能测试及演示，若测试结果没有达到供应商响应文件中承诺的技术要求，那么所产生的一切经济损失、后果和法律责任均由供应商承担。

2、以下打“★”项参数为必要条款，其中任何一条不满足视为无效报价；非打“★”项参数在不影响产品的功能特点和整体结构的基础上，允许适度偏离，但必须与询价文件所述相一致，偏离是否适度、是否科学合理，将由评委给予综合评价，供应商应承担被判定为偏离度过大，已影响产品功能特点和整体结构，报价无效的风险。

1. **软硬件产品技术参数要求**

**4.1数据中台软件产品**

**4.1.1数据开发平台**

数据开发是完成数据清洗、统计或数据挖掘的计算逻辑的一站式过程，它需要具备广泛的任务类型支持，强大的调度引擎和友好的团队协作式开发平台。提供可视化界面操作方式的大数据开发工具，包含数据同步，离线任务开发，任务管理等模块，提供任务开发和管理、脚本开发和管理、资源管理、函数管理等功能，具备实时任务调度和调度依赖功能，支持任务多版本管理。

|  |  |  |
| --- | --- | --- |
| 序号 | 功能模块 | 需求说明 |
| 1 | 配置数据源 | 支持多种异构数据源之间的数据读写，包括MySQL、Oracle、SQLServer、PostgreSQL、Hive、HDFS、ElasticSearch、HBase、Redis、MongoDB、FTP、Gbase等。支持CDH、FusionInsight、TiDB、GP、Oracle或EMR作为存储和计算引擎 |
| 2 | 数据同步 | 支持以可视化配置以及脚本模式等多种形式进行数据同步，支持数据同步速度配置；支持对关系型数据生成批量同步任务；支持配置同步规则，决定是否停止同步任务；★整库同步,分库分表同步：支持MySQL到Hive数据源的整库同步，以及分库分表同步★定点增量同步：支持对关系型数据库进行定点增量同步，下次同步时，可从上一次同步成功的点位继续开始运行支持同步任务运行状态监控与执行结果统计，包括读取记录数、读取字节数、同步速度、写入记录数、写入字节数等；★支持脏数据管理，支持对同步过程中异常的数据自动保存，或存储至指定位置；支持脏数据表管理，包括表的基本信息查看，支持脏数据产生趋势统计、脏数据产生原因分析；支持断点续传，系统可自动记录上次同步的数据点位，下次运行时，自动从上一次的位置继续同步；支持界面化的配置，实现实时采集MySQL，Oracle，SQLServer、日志类等数据源的日志数据。 |
| 3 | 数据开发 | 以可视化形式支持离线任务开发和流式任务开发，支持基于WEB的形式进行任务编写；支持丰富的任务类型，至少包括SparkSQL、数据同步、PySpark、原生Python、虚节点等多种任务类型，用户可构建各种类型的任务来进行数据处理；★支持工作流任务，通过托拉拽的方式快速配置任务间的依赖关系；支持表结构智能提示，根据用户编写的SQL代码自动识别出相关的表信息，辅助用户完成开发过程；SQL任务支持UDF函数管理；★支持任务版本记录，发布任务时，支持对SQL代码、同步任务脚本、任务调度、环境参数等各类信息做版本记录，并标注当前版本与历史版本的区别；支持系统参数，例如运行时间、当前月、当前年等，支持在SQL或数据同步等任务中使用系统参数，系统根据代码运行的时间自动解析；支持自定义参数，以多种时间基准进行定义，包含任务计划时间、任务运行时间、业务日期，满足数据开发多种数据处理需求；支持为每个任务配置不同的执行环境参数，例如内存、CPU核数等；调度与依赖配置：支持天、周、月、小时、分钟不同粒度的调度周期，支持多个任务间上下游依赖配置；资源管理：支持Jar包、Python等资源的上传、删除、替换；支持统一资源，不同用户上传的资源可以共享共用；支持单个任务跨周期的自依赖配置与任务调度，支持自依赖、跨周期依赖等多种模式；支持不同项目包含不同的人员，并内置管理员、数据开发等多种角色； |
| 4 | 运维管理 | 支持汇总统计当日运行的周期实例的状态，直观的掌握当天任务运行的情况支持按日统计实例的执行时长，并显示执行时间较长的任务，便于用户筛选大任务，进行针对性的代码优化支持数据重刷操作（补数据），支持选择业务日期时间段并批量生成相关实例运行，并进行补数据实例的运维，如任务重跑、任务杀死、置成功等支持统计周期实例的运行状态、运行时长、任务运行日志；查看任务DAG图，对上下游任务一目了然，并可逐级展开更多层级的任务；★支持任务的冻结、解冻操作，已冻结的任务禁止运行，且已冻结的任务才可删除；告警配置：支持对每个任务设置告警，支持按照失败/停止状态、定时未完成、超时未完成等状态触发告警 |
| 5 | 数据资产管理 | 支持对表的结构变更操作进行记录，包括表结构变更语句、操作用户、操作时间。支持表的元数据查看，包括表的创建者、创建时间、描述、存储量、生命周期、最近一次表结构变更时间、数据最后变更时间、分区信息等功能。★支持对表进行编辑，包括修改表名、新增字段、分区字段、所属类目、生命周期等。支持数据类目的构建，并将表指定到类目中；支持根据类目进行表查询。★支持表级、字段级血缘关系展现，支持沿血缘链路的逐层追溯。支持对数据表相关的任务（包括同步任务）、脚本的相关性分析，将相关任务/脚本列出。支持表级、字段级权限控制，项目空间内支持用户自由访问任意数据表，跨项目访问需提前经过审批授权。支持跨项目访问数据时的申请、审批、手动收回权限等操作，支持授权时效性控制，超出时效后权限自动回收。支持自定义脱敏规则，对某个字段进行脱敏，系统可根据血缘关系，将脱敏规则扩展至其他相关表对字段。支持对分区表执行数据生命周期管理，支持分区级生命周期管理，支持根据每个分区的数据最后变更时间刷新生命周期。支持对租户内的项目数量、表数量、占用存储等指标模型总览：支持对数据模型检测状态进行汇总统计，并统计模型、字段不规范的原因和产生趋势；检测中心：支持定期对数据模型（表名）进行规范性检测，包括模型层级、主题域、增量标识、刷新频率等参数进行检测，并输出不规范的原因；配置中心：支持模型层级、主题域、刷新频率、增量等多种参数的配置；支持对原子指标、衍生指标的新增、编辑、删除； |
| 6 | 项目管理 | ★支持双项目模式，即开发模式与生产模式，生产/开发环境的代码隔离、数据隔离，开发环境必须通过类似“发布”的动作，将代码传递到生产环境。 |

**4.1.2数据API服务**

通过API应用服务快速构建数据API，帮助提高数据开放和共享效率。

|  |  |  |
| --- | --- | --- |
| 序号 | 功能模块 | 需求说明 |
| 1 | API市场 | 数据源支持：支持配置MySQL、Oracle、SQLServer、PostgreSQL、kylin、Impala、阿里云ADS、MongDB、TIDB、ES存储等作为数据源；支持在API市场中查看API基础信息、订购调用情况、请求返回参数等，帮助API申请者判断该API的适用性，同时支持API申请的发起；支持从API管理者、API使用者不同的角度，整体掌握用户调用量TOP排行、接口调用量TOP排行、接口失败率TOP排行等；支持用户按照API调用次数申请，超出申请的次数时或超出时间段时，系统自动为此用户停止API服务。★支持向导模式和脚本模式生产API。 |
| 2 | API管理 | 支持外部API注册至API网关，支持post、get、put、delete请求，可进行API统一管理、发布、调用API调用情况：可查看最近24h、最近7天、最近30天、历史累计维度的调用次数统计、错误日志统计；★API订购情况统计：可查看该API订购用户数，每个订购用户的使用情况；支持管理者对所有用户或单独用户禁用某个API；★支持查看所有API申请、执行审批/拒绝操作；支持取消某用户的API授权。支持API文档导出，包含API基础信息、参数信息、请求示例等； |
| 3 | API调用 | ★支持AK/SK签名加密方式和API token方式调用，充分校验发送者的身份和保障数据传输的安全性；★支持设置黑白名单，控制API调用的IP；★支持行级权限控制，更细粒度控制数据调用权限★支持设置API的调用次数与调用周期，实现API调用的管理。 |

**4.1.3数据标签管理**

|  |  |  |
| --- | --- | --- |
| 序号 | 功能模块 | 需求说明 |
| 1 | 实体管理 | 支持实体创建，不同实体间标签相互隔离支持实体间通过相同字段进行匹配创建跨实体关系支持自定义字典快速配置，通过手动创建/上传以便原子标签等创建时快速映射中文 |
| 2 | 标签中心 | 支持原子标签快速映射支持基于原子标签的规则性衍生标签配置支持标签根据用户分组设置可见性发布，支持发布周期自动化管理支持最多三级类目管理 |
| 3 | 群组管理 | 支持通过配置多级规则圈选目标群组支持通过匹配主键根据本地表格文件圈选群组支持本地上传/标签圈群两种方式生成群组支持已有群组根据标签特征对比 |
| 4 | 标签市场 | 支持针对不同用户分组发布标签进行标签权限管理查看账户下已有标签权限，并进行管理操作支持高级用户对标签申请进行审批管理操作，支持手动收回标签权限； |
| 5 | 标签服务 | 支持将已有群组通过Token加密的API分享，支持输出群组时手动控制输出标签 |
| 6 | 项目管理 | 创建新的项目，相当于创建新的hive Database支持自定义用户分组，用于进行标签权限控制添加成员/移除成员/分配成员角色等 |

**4.1.4集群运维管理**

在产品的全生命周期中，部署升级是短时间节点性的操作，线上运行环境的稳定性保障非常重要，支持运维人员通过监控各个节点服务的运营状况，并在达到告警设置时通知告警接收人。

|  |  |  |
| --- | --- | --- |
| 序号 | 功能模块 | 需求说明 |
| 1 | 安装部署 | 图形化、流程化的部署向导，快速完成产品部署，实时查看服务部署进度及状态；可根据规则制作schema安装包，进行第三方产品部署监控，且产品安装包支持界面化接入；支持增量patch包，以轻量包升级产品，方便快速；★图形化快速配置服务与接入主机，且允许界面化修改参数；★支持引入外部服务，可直接使用集群外已部署的服务，不需要重新部署；支持服务继承，继承已部署的其它产品服务；支持查看历史部署记录以及历史部署快照，方便回顾历史部署情况，记录部署日志，方便查看安装过程错误。 |
| 2 | 集群运维 | 支持查看集群性能指标，如集群CPU使用率、内存使用率、I/O读写速率；支持查看各服务的运行状态、健康状态、性能指标、配置文件、安装日志；支持查看主机的状态、最近心跳时间、及性能指标；支持对各服务进行定时健康检查，及时查看健康检查状态；支持为服务进行扩容、缩容，支持产品版本快速升级和回滚；界面化修改各服务配置参数，修改完成后进行产品升级，服务自动更新；支持集群停止与启动、且支持断电回复后服务自动拉起；支持服务实例停止与启动、支持服务实例滚动重启、支持服务停止与启动。 |
| 3 | 多租户多集群 | ★支持多租户机制，不同租户之间的用户、数据保持隔离；支持按照租户分配集群计算资源，例如A租户分配30%资源，B租户分配20%资源，且多个租户之间在忙/闲时可以根据负载动态调整；支持一套开发平台（离线计算、实时计算）对接多个计算集群。 |
| 4 | 监控报警 | ★支持为产品及服务新增dashboard，并自行配置监控指标；支持dashboard的导入与导出；支持查看与配置各组件服务（如hive\hdfs\spark等）的多种性能指标；支持为组件服务的各指标配置告警；支持公有云短信、钉钉、邮件告警通道，同时可支持企业内部告警网关，满足专有云内部告警的需求； |

**4.2 硬件配置**

|  |  |  |
| --- | --- | --- |
| 项目 | 配置 | 数量 |
| 物理服务器 | 14\*3.5英寸SAS/SATA盘位直通功能SAS卡,冗余电源2\*CPU Intel® Xeon®Sliver 4210（10C，2.2GHz，14M）4\*32GB DDR4 RDIMMs 2933MT/s1\*240GB M.2 SATA SSD系统盘2\*千兆以太网口（RJ45)2\*万兆以太网SFP+光口（无模块）4\*32GB DDR4 RDIMMs 2933MT/s1\*240GB M.2 SATA SSD2\*960 GB SATA SSD 2.5英寸4\*8 TB 7.2K SATA 3.5英寸硬件三年7\*24技术支持服务 | 3 |
| 千兆交换机 | l 24 个 10/100/1000Base-T 以太网端口，4 个 100/1000 SFP，4 个千兆 SFP，2 个QSFP+堆叠口l 交流供电，电源前置，支持 RPS 冗余电源l 交换容量：598G/5.98Tbpsl 包转发率：168Mpps | 1 |
| 万兆交换机 | 24 个 10GE SFP+端口，6 个 40GE QSFP+端口l 可插拔双电源，支持 1+1 电源备份，默认不含电源l 4 个独立可插拔风扇，支持前后风道l 包转发率：720M/792Mppsl 交换容量：2.56T/25.6Tbps | 1 |

1. **系统技术要求**

大数据相关产品采用开源生态系统的基于Hadoop生态圈技术架构，包括平台关键技术组件、应用架构、平台引擎架构、性能设计、系统可用性设计、兼容性设计、开放性设计及安全性设计等。

数据应用统一门户必须支持分布式部署，建议采用SpringCloud 架构进行开发，前后端分离，系统服务之间要支持OAuth2.0协议进行授权。移动端支持和企业微信集成，在Android和iOS系统中显示操作正常。PC端支持WINDOWS、MACOS等操作系统，支持IE、360、Safari、等主流浏览器，支持大屏展示。

**5.1性能设计**

**（1）主要性能指标**

数据应用统一门户满足招标单位30000人的使用，支持平均在线用户量5000以上，支持2000个以上的在线用户并发访问。系统不能限制用户数，当用户量拥堵时可以横向扩展添加设备解决并发问题。

**（2）可用性指标**

保证7×24小时不间断稳定运行。

**（3）可靠性指标**

1）系统年度平均无故障率>＝99.5%。

2）系统应具备自动或手动恢复措施。

3）在充分备份情况下，系统意外宕机后的恢复时间不超过4小时。

4) 系统不存在单点故障。

5）系统有完善的备份体系以及全面的监控告警功能。

**（4）容错性指标**

1）集群中一台主机宕机，另一台主机仍然可以正常工作，对系统的性能没有大的影响。

2)系统提供报错保存进入日志，并支持根据情况控制日志输出详细程度。

**5.2系统可靠性设计**

**（1）主备切换**

服务器间一个主服务器出现问题时，自动或手动切换到备份服务器，切换时间不超过10分钟。服务器内部关键服务模块提供备份模块，当一个模块出现问题时，自动或手动切换到备份模块，切换时间不超过5分钟。

**（2）无故障运行**

系统应满足在每日、每月等关键周期的高低负载下运行的稳定性，保证系统在集中任务时间段内的系统性能要求，保证系统在一个月内宕机事件不超过一次。

系统应支持7\*24小时持续不间断的高稳定性运行。

系统每年重大故障累计时间不超过12小时。

**（3）故障发生时及时恢复**

系统应保证在发生系统故障时，1个小时内能恢复系统。系统应保证在系统集群设计的情况下，在发生单节点故障时系统接管的有效性。

**5.3安全设计**

大数据平台应当提供统一、规范的安全保证体系，主要从平台安全控制、安全风险控制、数据安全控制等三方面入手。

**（1）平台安全控制**

建立完善的账户认证安全体系，授予不同帐户完成各自承担任务所需的最小权限，他们之间的权限和资源相互隔离。

**（2）安全风险控制**

大数据平台建设时应充分考虑高可用，主从等风险管理机制，保证平台在故障发生时，能够继续提供一部分功能，确保能够实施必要的措施，并且应提供关键网络设备、通信线路和数据处理系统的硬件冗余，保证系统的可用性，能够对重要信息进行备份和恢复。

1）库、表、字段级别数据安全控制

对数据库表字段可以进行细粒度的权限控制，针对相应的表字段有细分到操作类型的权限控制。可对整库，整表，进行赋权。基于角色进行授权管理，针对不符合安全控制的执行请求进行相应拒绝。

2）对数据访问和计算进行管理审计

对用户数据操作与计算访问进行日志记录。可对操作进行安全审计评估。

3）数据传输过程安全权限控制

在数据传输过程中采取加密等措施控制数据安全。

1. **项目实施交付要求**

**6.1项目资料要求**

项目实施完成验收时，乙方应提供项目资料包括但不限于如下文档：

1.双方往来文件。

2.进度计划及完成情况。

3.项目人员情况（花名册、相关资料）。

4.项目需求文档/变更文档。

5.规划咨询文档

6.设计文档。

7.数据库设计文档。

8.接口文档。

9.测试文档（测试用例、测试报告）。

10.应用部署手册（服务器基本信息、部署环境版本信息、部署说明、备份恢复说明、密码修改、应用启停、服务端口脚本配置、参数配置、补丁升级更新说明）。

11.数据库部署手册（数据库基本信息、数据库脚本、备份恢复说明、参数配置、补丁升级更新说明）。

12.系统操作手册。

13.合同约定的其他文档和资料。所有的项目资料在终验前必须完整地移交给招标人。

**6.2项目保密要求**

未经甲方事先书面同意，乙方不得将由甲方或者代表甲方提供的有关合同或者任何合同条文、计划、图纸、模型或者资料提供给与履行本合同无关的任何其他人。即使向与履行本合同有关的人员提供，也应当注意保密并限于履行合同必须的范围。

乙方不得将甲方资源的配置、登录方式、资源账号密码，业务架构、系统负载规格等信息在任何时间、任何地点、任何方式提供给任何其它人。该信息即使乙方人员使用，也应当注意保密性。

1. **保修期及售后服务**

保修期范围：自软件完成验收通过之日起12个月。

人员要求：供应商在服务期内提供专职的服务人员提供服务，服务人员需具备云计算、大数据、网络、安全等专业知识和技能。

服务响应要求：供应商须在维保期内提供 24 小时全天候技术服务热线，负责解答用户在数据中台使用中遇到的问题，并及时提出解决问题的建议和操作方法，对故障15分钟内响应。

培训要求：现场培训，培训时间根据招标方后期要求确定，提供5人/天产品培训。

1. **项目完工期限**

采购人和成交单位应当自成交通知书发出之日起30日内，根据询价文件和成交单位的报价文件订立书面合同。采购人和成交单位不得订立背离合同实质性内容的其他协议。

成交单位无正当理由拒签合同的，采购人取消其成交资格。

1. **付款方式**

付款分为到货款、终验款两部分。其中，到货款占合同金额的80%，货到后支付。终验款占合同金额的20%，项目验收合格后支付（包括软件产品质量和技术服务质量）。