**数据高铁项目参数要求**

|  |  |  |
| --- | --- | --- |
| 序号 | 指标项 | 技术参数要求 |
| 1 | 基本功能要求 | 投标人需使用专业的数据实时同步复制软件工具完成医院数据实现采集，实现多库实时汇聚要求； |
| 提供服务所用数据实时同步复制软件须拥有自主知识产权，非封装OEM产品， |
| 支持通过分析数据库的日志文件对数据库中交易的实时抓取；支持日志并发分析，可通过一份分析结果同时对多个同步链路提供数据 |
| 2 |
| 3 | 本次需提供无限CPU数量及存储容量数据同步授权 |
| 4 | 性能要求 | 全量数据首次全同步效率不低于300GB/小时 |
| 5 | 增量数据实时同步效率每小时不低于200GB/小时 |
| 6 | 实时同步平均延时不超过5秒。 |
| 7 | 非侵入式安装要求 | 为保障生产安全，软件部署不得干预源端生产业务，禁止对源端数据库及操作系统做任何的参数修改，禁止重启服务器； |
| 8 | 软件必须尽可能少的占用系统资源，CPU占用不大于5%，内存占用不大于1G； |
| 9 | 支持中间机部署模式，无需在生产数据库服务器上安装代理程序 |
| 10 | 在数据初始化同步过程中，业务不能停止；在异构平台或跨数据库版本之间的数据初始化过程中不需要人工干预； |
| 11 | 与医疗机构对接仅限于数据库层面，不需要医疗机构做系统层面改造； |
| 12 | 异构硬件平台支持 | 软件支持任意硬件平台异构实时同步复制，此次需要支持windows操作系统到linux操作系统的数据同步； |
| 13 | 操作系统要求 | ★应支持以下操作系统：Linux主流品牌和版本（Redhat linux/CentOS/**中标麒麟/银河麒麟**等）/Windows/HP-UNIX 11.23 v2 for PA/HP-UNIX 11.23 v2 for IA/Sun Solaris 9,10/IBM AIX 5/AS400以上版本（提供证明材料） |
| 15 |  |  |
| 16 | 支持不同大数据组件（HDFS/HBASE/HIVE/KAFKA/ES等）、大数据平台（**CDH/华为/星环**等）的异构数据同步 |
| 17 | 针对不同数据库的部分非法数据，自动按照目标端规范调整适配； |
| 18 | 支持对数据源进行分类管理； |
| 19 | ★支持从产品/项目的维度对数据源进行授权，授权后相应的产品/项目才能使用该数据源（提供证明材料）； |
| 20 | 支持对数据源中心的核心指标进行统计、如数据源总数、连接状态等； |
| 21 | 对数据源的连接状态进行监控，并支持查看数据源产生的告警信息； |
| 22 | 支持源端字段变化感知以及产生告警，支持查看相应的告警信息。 |
| 23 | 元数据中心要求 | ★支持用户自定义元模型，通过关联元模型进行采集，实现动态维护元数据表中的管理类属性和业务类属性；（提供证明材料） |
| 24 | 支持为采集任务配置定时调度策略； |
| 25 | 支持查看采集任务执行的作业信息，通过查看日志，可查看本次采集时新增/删除的表、视图和字段； |
| 26 | 支持通过系统、数据源目录的方式展示元数据信息，包括表元数据、字段元数据、视图、存储过程等元数据信息； |
| 27 | ★支持自动生成数据库存储过程的字段级血缘关系，支持血缘关系图谱的放大缩小和导出成图片；（提供证明材料） |
| 28 | 支持表级、字段级的血缘跟踪与展示，类型包含：血统分析、影响分析和全链分析； |
| 29 | 支持根据采集的时间，进行历史数据比对； |
| 30 | 支持建立关联数据，实现数据的下钻查询； |
| 31 | 数据集成 | ★支持通过拖拉拽的方式，只需要拖动输入组件、转换组件、输出组件、配置组件连线等方式就可以快速完成定时数据同步任务的配置；（提供证明材料） |
| ★支持提交新的版本，并通过版本回滚的方式回退版本。支持对同一个任务的两个版本进行比对，并以高亮的方式显示基础信息、任务配置、组件信息等差异数据。支持对提交版本进行任务配置信息检查、表重复性检查、各个ETL组件配置风险的自动化检查；（提供证明材料） |
| 32 | 支持多表Join、字段级Trans、数据过滤、字段映射、去重、排序、UTF函数等多种数据转换功能。支持同一任务内不同源库的表通过union后同步至目标端单表或者单文件。支持输入/输出组件前置SQL和后置SQL的配置； |
| 33 | ★支持把已经开发好的定时同步任务，通过调度任务进行串联或者并联的工作流配置，支持定时任务配置，支持对工作流实例进行重跑、中止、强制成功等操作；支持工作流详细运行结果以及日志的查看；（提供证明材料） |
| 34 | 支持对不同的调度工作流进行调度依赖配置，依赖配置组件支持配置多个工作流、多个任务的“且”和“或”的多重依赖，并根据工作流依赖形成工作流关系； |
| 35 | 支持对提交版本的定时离线任务进行任务发布、定时配置、查看编排、执行一次、执行记录进行操作运维，能够对任务的上线启用情况、最近一次执行结果、任务状态等做统一查看； |
| 36 | 支持对所有运行的任务实例进行概览统计，包括运行次数，成功次数、失败次数、抽取总量、加载总量、上线天数等信息； |
| 37 | 支持对平台自有任务的导出和导入功能，同时可以兼容主流的数据ETL产品，如DataStage、PowerCenter等产品任务XMl文件自动化的导入、任务解析、任务自动生成，任务生成后可以完整的复原导入任务的各个ETL组件、组件内的表结构、数据转换逻辑、数据映射关系、字段级元数据信息等； |
| 38 | 支持对同步任务端到端的数据表级、字段级的血缘分析和展示，可以有效的支持字段级数据质量的定位和任务逻辑错误的修复；  |
| 39 | 任务调度中心 | 支持为任务配置定时调度策略，策略有：每日策略、每周策略、每月策略和自定义Cron表达式。 |
| 40 | 支持对任务运行的实例进行集中管理，可查看其运行状态、日志详情等； |
| 41 | 支持对任务进行工作流的配置，通过DAG拖拽的方式完成任务工作流配置，实现任务的定时调度、顺序执行。 |
| 42 | 数据库实时抽取要求 | 支持一对一，双向，一对多，多对一，和级联复制 |
| 43 | 支持对作业进行运行、停止、恢复操作；支持检查作业的依赖关系，可通过资源管理器的资源分配将作业分发到调度代理执行作业 |
| 44 | 支持数据转换、数据拆分及分发； |
| 45 | 支持从多个数据库中同一类型表的记录整合到一个表中去； |
| 46 | 支持数据转换，包括列映射、增/删除列、列转换； |
| 47 | 支持DML操作复制、支持SEQUENCE、函数、存储过程、视图、同义词、索引、应用包、用户等数据库对象进行复制 |
| 48 | 支持没有PK/UK字段的表的复制、并无需打开或修改数据库参数，可定义并过滤不需要复制的事务； |
| 49 | 支持按照schema方式设置复制关系，无需单表设置复制关系，支持不同源和目标端在不同的schema名情况下的复制； |
| 50 | 支持中文汉字内码，符合双字6节编码；支持DXF数据格式的装载；支持Rowid mapping的方式实现数据快速定位； |
| 51 | ★支持从数据库日志中分析所有DDL操作，不影响整个同步过程。在源端数据库数据结构发生变化（如修改表结构，增加表等）后，复制软件可自动将新增结构变化同步到目标端，支持DDL策略配置。（提供证明材料） |
| 52 | 支持如下转换操作 |
| 列 操 作： 增加列、过滤列、变更列 |
| 行 操 作： 条件过滤选取 |
| 数据操作： 函数变换、类型转换、合并、算术运算、几何运算、乘数运算 |
| 分区操作： 分区抽取、分区复制、表变分区、分区变表 |
| 结构操作： 多表合并、一表分裂、多层组合 |
| 字符操作： 多种字符集整理成一种字符集 |
| 跨表转化：参照外部表进行数据转换 |
| 53 | 数据库同步复制要求 | 在同步sqlserver数据库过程中，目标数据库须始终处于可用状态，数据同步过程中，支持关闭CDC代理作业，避免对生产库性能造成影响；  |
| 54 | 软件需支持数据库表、视图、函数、存储过程、序列等对象的实时同步。且源端表结构修改后（增加删除列），无需对表进行重新初始化同步，仍能正确实时同步表数据（DML）； |
| 55 | 软件支持sqlserver数据库双向同步； |
| 56 | 软件支持跨平台部署，既可以部署在windows主机，也可以部署在linux主机； |
| 57 | 软件支持alwsys on模式部署，软件能够做到全量和增量主要分析辅助节点的数据，仅初始化开启附加日志、增量DDL分析、增量LOB回查、增量日志链断裂访问日志备份文件时访问主节点，避免影响生产业务。 |
| 58 | 安全认证支持 | 支持Oracle数据实时复制到基于kerbors安全认证的kafka环境，支持json格式数据实时入kerbors安全认证的kafka，支持可自定义包含json格式模板数据实时入kerbors安全认证的kafka； |
| 59 | 复杂环境应对能力 | 支持无主键表同步 |
| 支持“image/text/ntext”大字段数据类型的实时同步 |
| 支持目标端自定义指定主键列 |
| 支持过滤一些自增约束 |
| 60 | 数据传输安全要求 | 数据传输过程中采用严格加密方式防止数据泄漏 |
| 61 | 支持文件加密传输，秘钥长度不小于256位 |
| 62 | 非结构化数据同步 | 支持阿里OSS、华为OBS、新华三、星辰天合、日立、青云、腾讯COS对象存储同步功能。 |
| 63 | 支持NAS、NFS、文件系统文件同步功能。 |
| 64 | 异构数据库同步支持 | 软件支持源端Caché数据库到PostgreSQL异构数据库环境下的数据实时复制（两端延时在5秒以内），且目标端PostgreSQL数据库可实时查询，当生产端插入（Insert/ Update/delete）数据后，目标端PostgreSQL数据库无需启动，始终处于实时打开验证状态，并通过数据库select语句在5秒以内，查询到该数据变化。 |
| 提供此项功能截图并加盖供应商公章的证明文件，截图内容为真实环境或者自己搭建模拟环境，需展现源端、目标端IP地址、数据变化轨迹以及软件运行日志**；** |
| 65 | 关键表保护功能 | 软件需支持关键表保护功能： |
| （1）源端对某张关键表执行drop 操作；（2）查询目标端，这张表已经同步进行drop操作，查询无此表或者无此数据；（3）查询目标端被删除的表，能够备份成另一张表，表的内容和误删的表一致。 |
| 66 | 数据比对功能 | 支持数据比对功能，支持指定不同的数据比对策略，如全库比对、指定表比对、指定列比对；可根据需求制定比对计划和比对任务，对源端和目标端的数据进行比对和校验 |
| 67 | 故障应对能力 | 在遇到系统错误引起的复制中断时，例如硬件故障、数据库故障、网络中断或延迟，分级存储机制能完好的保存已经合成的交易信息，避免数据丢失，直到系统故障解决，恢复从队列传输的中断点开始； |
| 68 | 数据回溯 | 可通过日志分析记录源端数据库数据内容/操作类型（插入/删除/更新）/操作时间；可记录重点数据行列数操作类型/操作内容 |
| 69 | 采集数据管理功能 | 支持把采集后的数据表发布到对应的目录下，形成大数据服务门户（库表、接口、文件数据服务方式；支持采集数据资产管理功能，具备数据查询检索、数据申请订阅、数据共享交换功能，同时用于指导后续按照主题域划分逐步进行数据治理**；** |
| 70 | 自主知识产权要求 | 为避免使用的软件产品是国外产品封装或者开源产品简单封装而成，避免不必要的安全漏洞风险，软件核心模块的开源代码占比不得超过20%  |
| 72 | 支持故障诊断及告警 | 提供具有故障的监视和诊断能力，平台需支持主动告警功能，出现不正常情况时，能够第一时间主动掌握目标数据库及汇聚链路异常状态，需支持邮件、短信、微信、钉钉告警方式，缺一不可 |
| 73 | 同步链路状态监控 | 监控平台需支持同步链路涉及的源数据库和目标数据库状态监控，如会话、连接数、表空间使用率、数据同步延迟等。 |