

团 体 标 准

T/ACEF XXXX—2024

生态环境大数据平台数据接入技术规范

Technical specification for data access of ecological environment big data platform

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

2024-XX-XX 发布

2024-XX-XX 实施

中华环保联合会 发布

目 次

前 言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 接入范围	1
5.1 数据源	1
5.2 生态环境基础信息资源	2
5.3 主题域数据资源	3
5.4 行业相关数据资源	3
5.5 互联网数据资源	3
6 接入方式	3
6.1 关系数据库抽取	4
6.2 数据网关	5
6.3 文件接收 FTP 服务	5
6.4 文件拉取 FTP 服务	7
6.5 消息队列	8
6.6 文件 HTTP 服务	9
6.7 文件 NFS 服务	11
6.8 区块链数据服务	13
7 接入质量管理	14
8 接入安全管理	14
8.1 数据源认证	14
8.2 数据可靠传输	14
附 录 A （资料性） 智能质控生态环境物联网设备接入说明	15
A.1 申请数字证书	15
A.2 设备接入	15
A.3 数据传输	15
A.4 接口上传数据格式	15
参 考 文 献	16

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由北京环丁环保大数据研究院提出。

本文件由中华环保联合会归口。

本文件起草单位：北京环丁环保大数据研究院、北京市生态环境监测中心、北京微芯区块链与边缘计算研究院、联通数字科技有限公司。

本文件主要起草人：…、…、…。

生态环境大数据平台数据接入技术规范

1 范围

本文件规定了生态环境大数据平台数据的接入范围、接入方式、接入质量管理、接入安全管理等方面的要求。

本文件适用于生态环境大数据平台数据接入。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

HJ 718 环境信息共享互联互通平台总体框架技术规范

3 术语和定义

HJ 718-2014界定的以及下列术语和定义适用于本文件。

3.1

生态环境大数据 ecological and environmental big data

具有体量巨大、来源多样、流动高速、形态多变等特征，并难以用传统数据体系结构有效处理的生态环境数据集。

3.2

接口 data interface

两个功能单元之间的共享边界，该边界由两个功能单元的功能特性、物理互连特性、信号交换特性及其他适当特性定义。

【来源：HJ 718-2014，3.2】

3.3

结构化数据 structured data

数据结构固定，能够用统一的结构表示的数据。本规范指存储于业务系统关系数据库中的数据。

3.4

非结构化数据 unstructured data

数据结构不固定，且不能够用统一的结构表示的数据。包括所有格式的办公文档、文本、图片、XML、HTML、各类报表、图像、音频和视频信息等数据。

4 缩略语

下列缩略语适用于本文件。

FTP: 文件传输协议 (File Transfer Protocol)。

HTTP: 标准的超文件传输协议 (Hyper Text Transfer Protocol)。

NFS: 网络文件系统 (Network File System)。

5 接入范围

5.1 数据源

接入的数据源包括：生态环境基础信息资源（5.2）、主题域数据资源（5.3）、行业相关数据资源（5.4）。

5.2 生态环境基础信息资源

生态环境基础信息数据资源见表1：

表 1 生态环境基础信息数据资源

一级类目	二级类目	三级类目
生态环境基础信息资源	环境质量	水环境质量
		大气环境质量
		土壤环境质量
		自然生态环境质量
		声环境质量
		海洋生态环境质量
		其他
	污染源	废水污染源
		废气污染源
		固体废物企业
		重金属污染源
		挥发性有机物污染源
		环境应急风险源
		交通运输污染源
		集中式污染治理设施
		施工工地污染源
		生活污染源
		农业污染源
		其他
		环境管理
	生态保护	
	应对气候变化	
	固体废物与化学品	
	核与辐射安全管理	
	环境安全应急管理	
	环境舆情管理	

表 1 生态环境基础信息数据资源（续）

一级类目	二级类目	三级类目
		环境行政许可管理
		生态环境监测
		生态环境保护督察
		环境执法
		其他
	其他	环境行政办公
		环境综合管理
		环境科技
		政策法规
		标准规范
		人事管理
		财务与审计
		环境地理信息
		其他

5.3 主题域数据资源

主题域数据资源包括水环境、大气环境、土壤环境、海洋生态环境、自然生态、固废与危化品、核与辐射、声、应对气候变化、行政许可、生态环境监测、环境督察、环境执法、环境统计、环境应急、环境舆情管理等。

5.4 行业相关数据资源

从其他部门接入的与环境相关的数据包括但不限于：

- a) 住房和城乡建设部门的工地数据和污水处理厂数据；
- b) 公安部门的机动车数据；
- c) 发展和改革委员会的燃煤数据和项目批复建设数据；
- d) 水利部门的水文数据；
- e) 自然资源部门的地下水、地质、矿权数据；
- f) 农业农村部门的土壤相关数据；
- g) 气象部门的大气相关数据；
- h) 海洋部门的功能区划和入海水源数据等。

5.5 互联网数据资源

在互联网上与生态环境相关的网络数据，包括包括但不限于：

- a) 生态环境部门网站；
- b) 生态环境系统网站；
- c) 生态环境媒体网站；
- d) 生态环境科技期刊网站、微信微博、门户网站等。

6 接入方式

6.1 关系数据库抽取

6.1.1 功能要求

关系数据库数据抽取应具备以下主要功能：

- 支持对主流的关系数据库进行数据抽取；支持对数据库中常用的数据类型进行数据抽取，至少包括数值型、字符型、日期/时间型等数据类型；
- 支持“全量”和“增量”两种数据抽取模式；
- 支持关系数据库中结构化数据抽取到大数据平台，包含结构化数据、半结构化数据存储的数据仓库中；
- 支持对关系数据库数据的采集内容和类型转换操作，至少包括选择具体的数据表、选择表中具体的字段、字段类型格式转换等操作；
- 支持数据抽取操作的立即执行、定期、不定期调度运行。定期调度运行应提供多种调度策略，至少包括固定期间间隔运行、指定期间点运行、指定期间范围运行、一次或指定次数运行等策略；
- 应提供图形化管理界面，包括数据抽取模式设置、抽取源关系数据库配置、指定数据表配置、表字段选择配置、字段类型转换配置、大数据平台目标存储位置配置、运行策略配置等操作界面；
- 应提供完善的日志和审计能力，可以记录数据抽取操作配置、运行时发生的各种事件；
- 应提供完善的监控机制，运行过程中出现异常可快速的定位及解决。

6.1.2 场景描述

应用场景描述如下：

- 关系数据库抽取服务，抽取数据源中数据库的元数据信息，包括数据库用户所属的表、字段信息；
- 关系数据库抽取服务配置数据传输任务，包括抽取数据库的源表和大数据平台对应的存储目标表；
- 关系数据库抽取服务配置数据传输任务运行策略，包括运行的开始时间、结束时间、运行频率；
- 关系数据库抽取服务运行数据传输任务，从数据源的数据库抽取数据到平台数据存储中。

6.1.3 应用场景图

关系数据库抽取应用场景见图1。

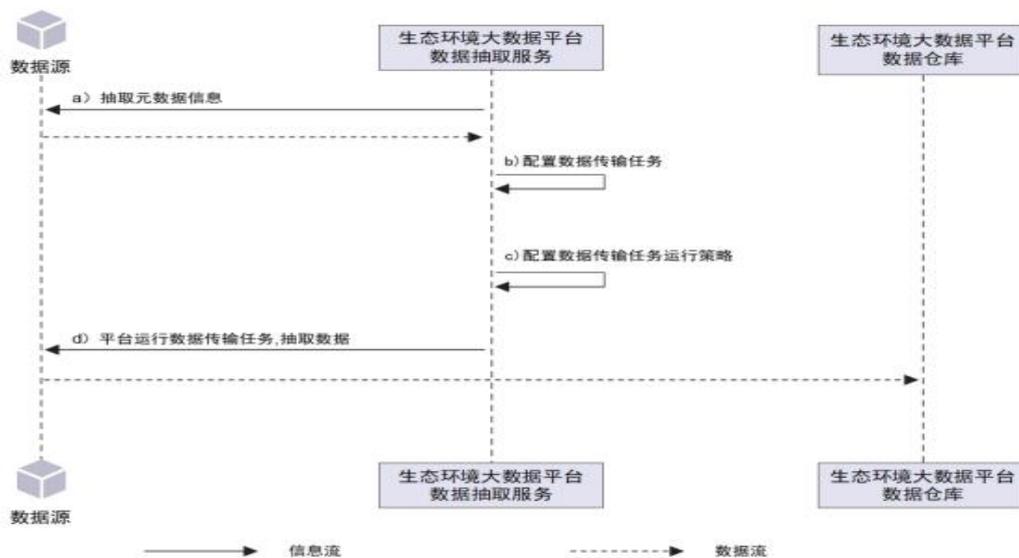


图 1 关系数据库抽取应用场景

6.2 数据网关

6.2.1 功能要求

数据网关服务为数据源提供大数据平台中结构化数据或非结构化数据的接口数据接入。网关服务应具备以下主要功能：

- 支持接入 webservice、RESTful 方式的接口；
- 支持包括结构化数据、非结构化数据的接口；
- 支持接口编排，轻松实现多个接口的功能集成；
- 提供图形化管理界面，用于接口数据存储位置、操作用户、目标存储位置的配置；
- 提供完善的日志和审计能力，应记录接口数据配置及数据抽取操作配置、运行时发生的各种事件；
- 具备熔断管理机制，接口访问异常情况下的处理策略，保证服务整体可用。

6.2.2 应用场景

数据网关服务应用场景描述如下：

- 数据源向大数据平台提供接口信息，包括：接口访问地址、输入参数、输出参数、验证方式等接口信息；
- 大数据平台根据数据源提供的数据接口进行定义及编排；
- 网关服务配置数据传输任务运行策略，包括运行的开始时间、结束时间、运行频度；
- 网关服务运行数据传输任务，从数据源数据接口中抽取数据到大数据平台数据仓库中。

6.2.3 应用场景图

数据网关服务应用场景见图2。

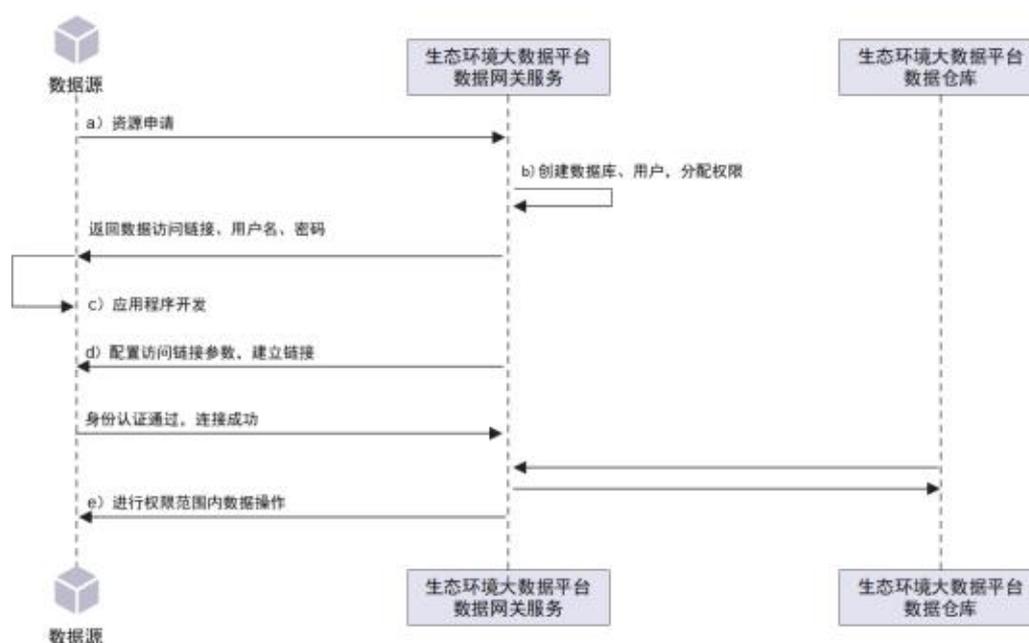


图 2 数据网关服务应用场景

6.3 文件接收 FTP 服务

6.3.1 功能要求

文件接收FTP服务应提供外部系统文件数据的接收并存入到大数据平台数据仓库的功能。文件采集应具备以下主要功能：

- a) 支持标准 FTP 协议接收数据；
- b) 支持顺序型断点续传功能；
- c) 支持接收的文件的重命名及指定存储目录；
- d) 应支持对接收文件的完整性校验；
- e) 应支持对客户端进行认证；
- f) 支持图形管理功能，支持认证配置、文件目标位置配置、校验处理配置。

6.3.2 应用场景

应用场景描述如下：

- a) 大数据平台配置应用账号、接收参数及存储位置；
- b) 管理信息系统通过标准 FTP 协议连接大数据平台服务；
- c) 管理信息系统检查目标临时文件是否存在；
- d) 管理信息系统发起全量或续传指令；
- e) 管理信息系统发送数据到大数据平台；
- f) 大数据平台接收文件数据；
- g) 管理信息系统发送数据校验文件；
- h) 大数据平台根据校验文件校验数据文件内容；
- i) 大数据平台按配置的规则存储接收到的数据文件；
- j) 大数据平台回写数据存储状态；
- k) 管理信息系统获取数据存储状态。

6.3.3 应用场景图

文件接收FTP服务应用场景见图3。

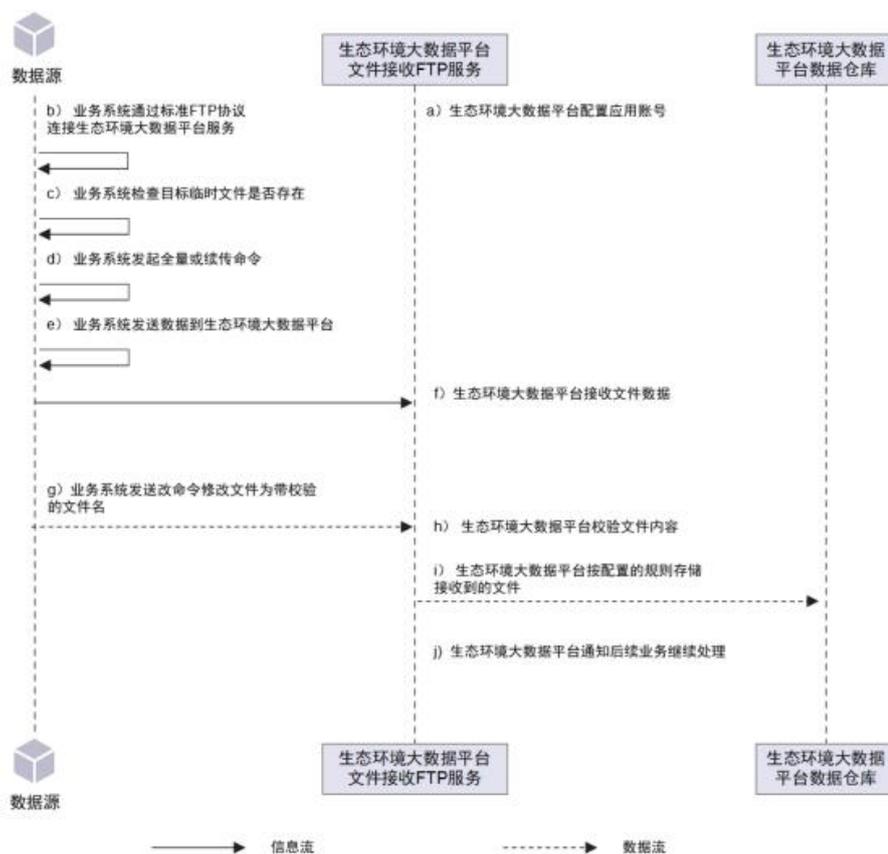


图 3 文件接收 FTP 服务应用场景

6.4 文件拉取 FTP 服务

6.4.1 功能要求

文件拉取FTP服务，应提供通过访问FTP协议实现将文件数据抽取到大数据平台数据仓库的功能。文件拉取FTP服务应具备以下主要功能：

- a) 支持顺序型断点续传功能，支持外部文件存储断点续传能力的自动识别及模式匹配；
- b) 支持 FTP 服务登录用户名和密码设置；
- c) 支持文件压缩传输，提供文件压缩规则设置；
- d) 支持文件加密传输，提供文件加密传输规则设置；
- e) 支持设置文件同步、异步拉取，支持设置拉取并行度；
- f) 支持指定目标文件存储位置、文件名，提供文件类型转换规则，支持常见类型转换；
- g) 支持全量文件采集，支持外部数据一次性初始化导入；
- h) 支持定期轮询文件采集，采集新增的文件，支持文件列表规则过滤；
- i) 支持图形管理功能，支持 FTP 连接配置、文件源配置、文件目标存储配置、文件压缩和加密传输规则配置、文件同步/异步传输规则配置、文件传输并行度配置、文件定期及实时策略配置、文件采集过滤配置。

6.4.2 应用场景

6.4.2.1 基于 FTP 协议的全量文件采集应用场景

应用场景描述如下：

- a) 业务系统应提供初始文件存储位置；
- b) 大数据平台全量文件采集提供基于 FTP 协议的采集任务配置，包括：存储文件位置、采集文件列表方式、是否文件校验、存储目标位置及文件存储命名规则；
- c) 大数据平台运行文件采集任务，读取文件存入大数据平台数据仓库中。

6.4.2.2 基于 FTP 协议的定期轮询采集应用场景

应用场景描述如下：

- a) 业务系统应按照业务需求生成数据文件，存储在指定文件区域；
- b) 业务系统应根据校验算法 SHA1 在相同目录下生成校验文件，校验文件名称与数据文件相同；
- c) 大数据平台轮询文件采集服务配置基于 FTP 协议的定期轮询采集任务，定期轮询采集文件；
- d) 大数据平台定期读取解析索引及校验文件内容；如果业务系统数据文件出现新增时，大数据平台读取数据文件内容，并且进行文件校验，如果文件完整，则将数据文件写入大数据平台中。

6.4.3 应用场景图

基于 FTP 协议的全量文件采集应用场景见图4，基于 FTP 协议的定期轮询采集应用场景见图5。

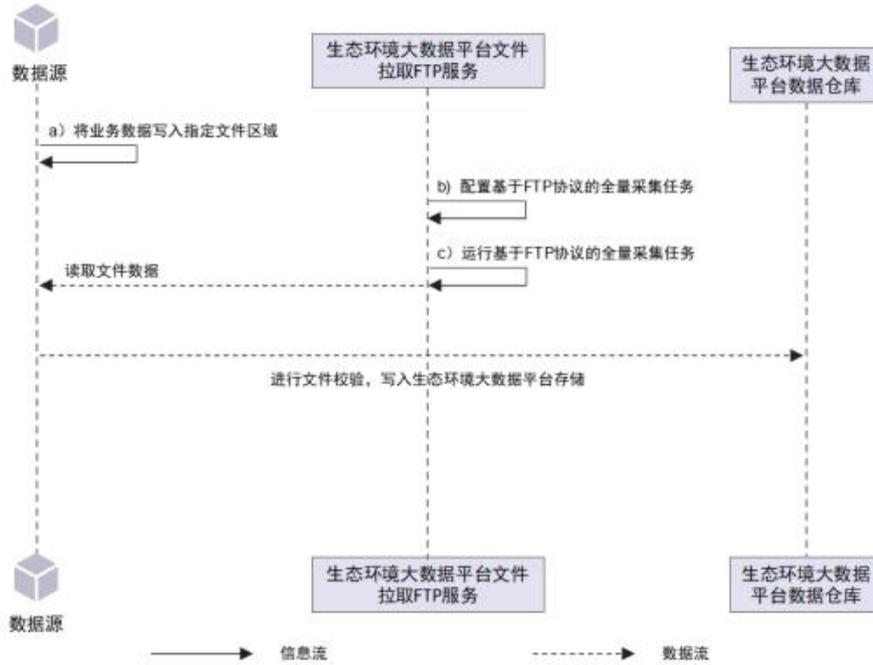


图 4 基于 FTP 协议的全量文件采集应用场景

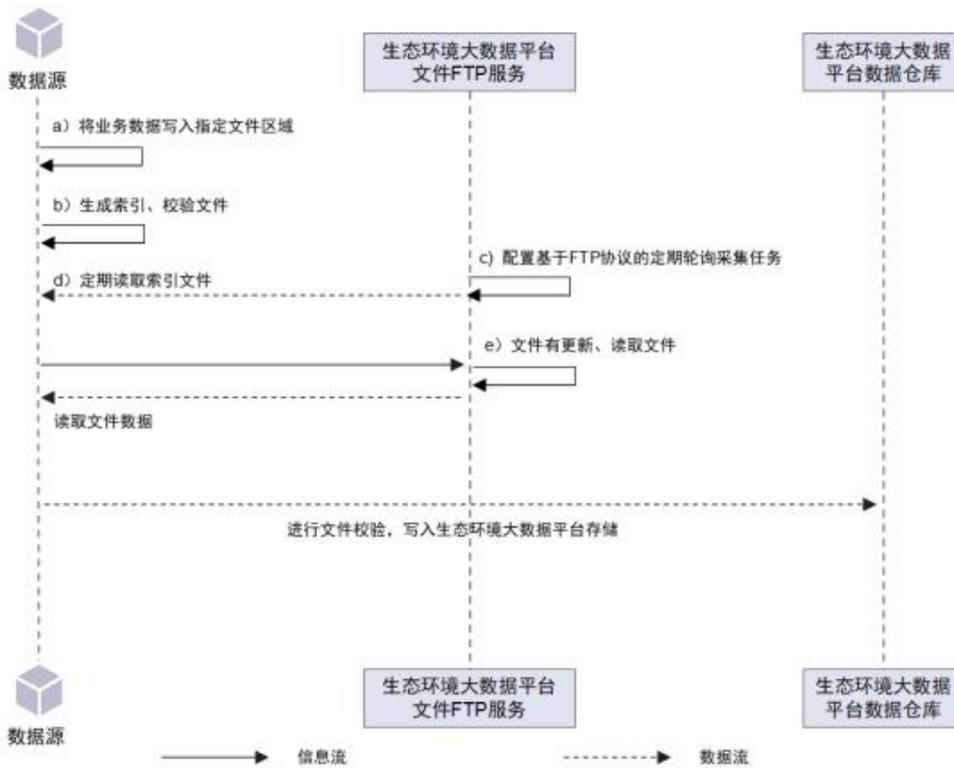


图 5 基于 FTP 协议的定期轮询采集应用场景

6.5 消息队列

6.5.1 功能要求

消息队列采集为管理信息系统提供消息推送和缓存功能。消息队列应具备以下主要功能：

- 应提供分布式消息队列的管理功能，支持消息主题的创建、删除、修改；
- 应提供支持“点对点”和“发布-订阅”两个消息模式；
- 应支持消息的持久化存储操作并且支持持久化周期设置；
- 应提供消息分布式高可用的发送和消费接口，包括链接建立、消息发送、消息消费、链接关闭，支持消息分区和备份操作；
- 具有风格统一的图形化管理界面，支持消息队列主题的创建、删除、测试、授权访问的操作；
- 具备完善的日志审计能力，应记录消息发送和消费时发生的各种事件。

6.5.2 应用场景

应用场景描述如下：

- 管理信息系统应向大数据平台申请消息队列接入服务；
- 大数据平台根据申请创建消息队列主题，返回消息队列名称；
- 管理信息系统开发业务处理程序，调用平台消息队列接口，发送数据或接收数据。

6.5.3 应用场景图

消息队列应用场景见图 6。

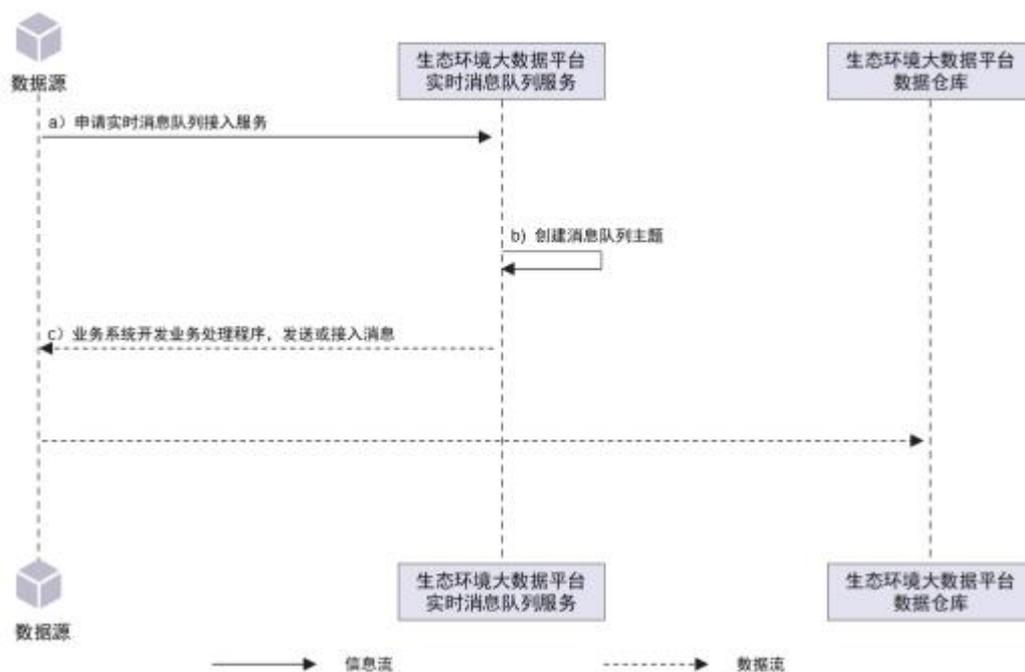


图 6 消息队列应用场景

6.6 文件 HTTP 服务

6.6.1 功能要求

文件HTTP服务，应提供通过访问HTTP协议实现将文件信息抽取到大数据平台存储的功能。文件HTTP服务应具备以下主要功能：

- 支持顺序型断点续传功能，支持外部文件存储断点续传能力的自动识别及模式匹配；
- 支持文件压缩传输，提供文件压缩传输规则设置；
- 支持文件加密传输，提供文件加密传输规则设置；

- d) 支持设置文件同步、异步拉取，支持设置拉取并行度；
- e) 支持制定目标文件存储位置、文件名，提供文件类型转换规则，支持常见类型转换；
- f) 支持全量文件采集，支持外部数据一次性初始化导入；
- g) 支持定期轮询文件采集，采集新增的文件，支持文件列表规则过滤；
- h) 支持图形管理功能，支持文件源配置、文件目标存储配置、文件压缩和加密传输规则配置、文件同步/异步传输规则配置、文件传输并行度配置、文件定期及实时策略配置、文件采集过滤配置。

6.6.2 应用场景

6.6.2.1 基于 HTTP 协议的全量采集应用场景

应用场景描述如下：

- a) 业务系统应提供初始文件存储位置；
- b) 大数据平台全量文件采集提供基于 HTTP 协议的采集任务配置，包括：存储文件位置、采集文件列表方式、是否文件校验、传输是否压缩、加密、同步/异步、并行度、存储目标位置及文件存储类型转换和命名规则；
- c) 大数据平台运行文件采集任务，读取文件存入大数据平台数据仓库中。

6.6.2.2 基于 HTTP 协议的定期轮询采集应用场景

应用场景描述如下：

- a) 业务系统应按照业务需求生成数据文件，存储在指定文件区域；
- b) 业务系统应根据校验算法 SHA1 在相同目录下生成校验文件，校验文件名称与数据文件相同；
- c) 大数据平台轮询文件采集服务配置基于 HTTP 协议的定期轮询采集任务，定期轮询采集文件；
- d) 大数据平台定期读取解析索引及校验文件内容；如果业务系统数据文件出现新增时，大数据平台读取数据文件内容，并且进行文件校验，如果文件完整，则将数据文件写入大数据平台中。

6.6.3 应用场景图

基于 HTTP 协议的全量采集应用场景见图7，基于 HTTP 协议的定期轮询采集应用场景见图8。

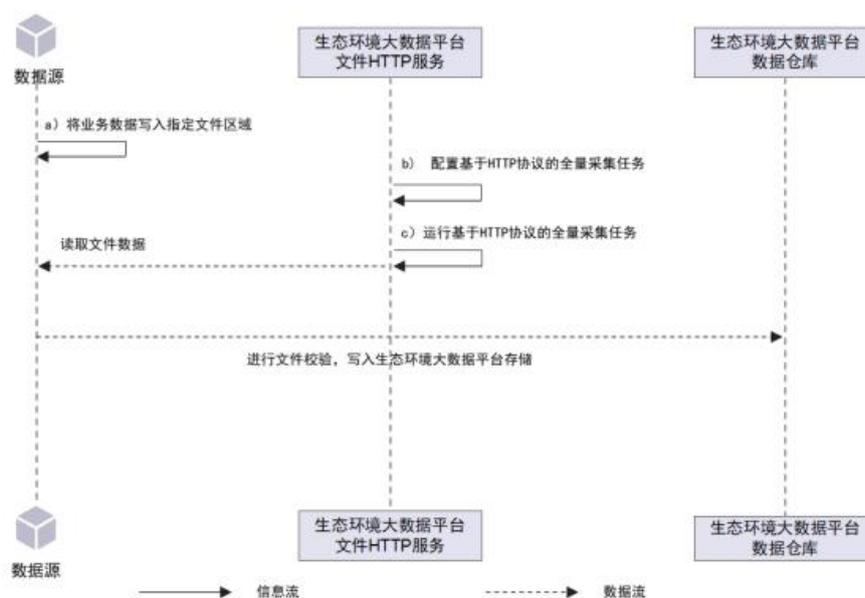


图7 基于 HTTP 协议的全量采集应用场景

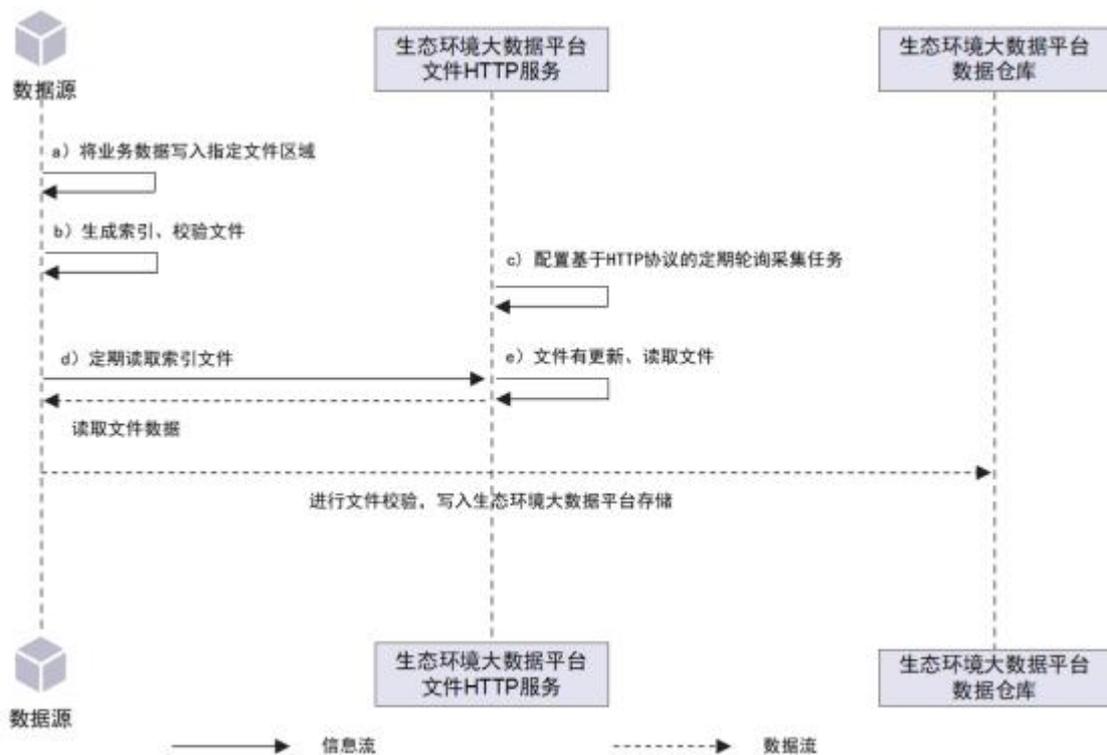


图 8 基于 HTTP 协议的定期轮询采集应用场景

6.7 文件 NFS 服务

6.7.1 功能要求

文件NFS服务，应提供通过访问NFS文件系统，实现将文件信息抽取到大数据平台数据存储的功能。文件NFS服务应具备以下主要功能：

- 支持顺序型断点续传功能，支持外部文件存储断点续传能力的自动识别及模式匹配；
- 支持指定目标文件存储位置、文件名，提供文件类型转换规则，支持常见类型转换；
- 支持全量文件采集，支持外部数据一次性初始化导入；
- 支持定期轮询文件采集，采集新增的文件，支持文件列表规则过滤；
- 支持图形管理功能，支持文件源配置、文件目标存储配置、文件压缩和加密传输规则配置、文件定期及实时策略配置、文件采集过滤配置。

6.7.2 应用场景

6.7.2.1 NFS 全量文件采集应用场景

应用场景描述如下：

- 业务系统应提供初始文件存储位置；
- 大数据平台全量文件采集提供采集 NFS 文件任务配置，包括：存储文件位置、采集文件列表方式、是否文件校验、存储目标位置及文件存储命名规则；
- 大数据平台运行文件采集任务，读取文件存入大数据平台数据仓库中。

6.7.2.2 NFS 文件定期轮询采集应用场景

应用场景描述如下：

- 业务系统应按照业务需求生成数据文件，存储在指定文件区域；

- b) 业务系统应根据校验算法 SM3 在相同目录下生成校验文件，校验文件名称与数据文件相同，文件扩展名为.SM3；
- c) 大数据平台轮询文件采集服务配置 NFS 文件定期轮询采集任务，定期轮询采集文件；
- d) 大数据平台定期读取解析索引及校验文件内容；如果业务系统数据文件出现新增时，大数据平台读取数据文件内容，并且进行文件校验，如果文件完整，则将数据文件写入大数据平台中。

6.7.3 应用场景图

NFS 全量文件采集应用场景见图9，NFS 文件定期轮询采集应用场景见图10。

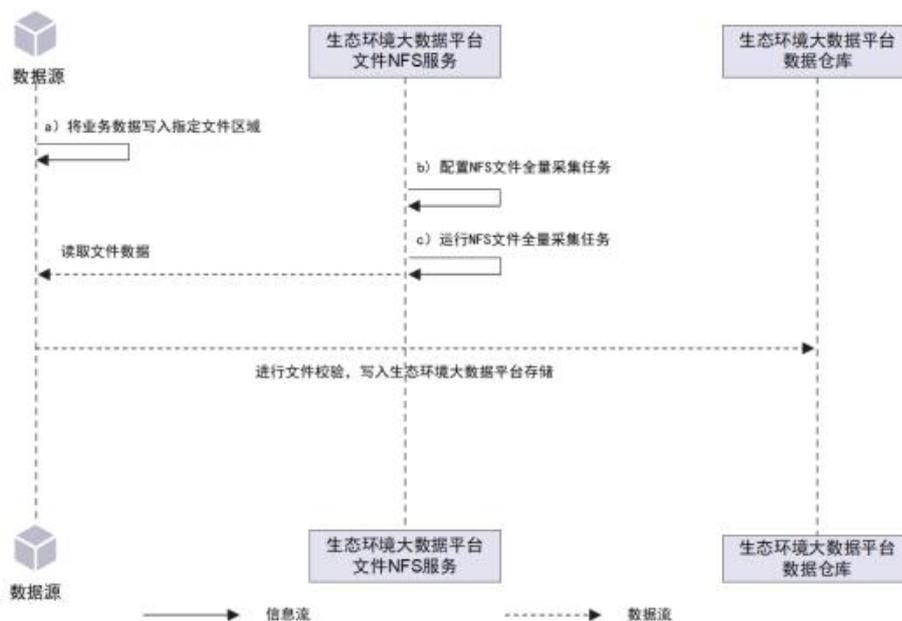


图 9 NFS 全量文件采集应用场景

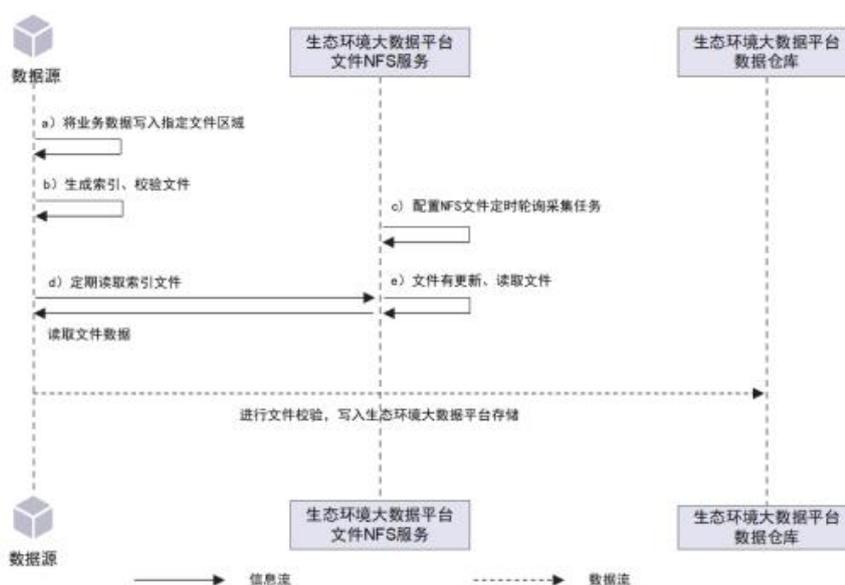


图 10 NFS 文件定期轮询采集应用场景

6.8 区块链数据服务

6.8.1 功能要求

生态环境大数据平台应具备区块链双链系统，分别是数据目录链和数据链系统。环境监测设备或环境应用系统的数据通过目录链公示可上链或应上链数据，通过代理节点或自身实现数据存储到数据链。区块链数据服务应具有以下功能：

- 接收环境监测设备编号信息并进行证书申请，进行设备的认证认可；
- 区块链节点应支持环境监测设备通过代理实现数据上链；
- 支持以全量表、快照表、拉链表 3 种方式进行数据存证版本确认；
- 支持上链数据的存储与检索；
- 支持上链数据的可控共享与交换；
- 支持上链数据的使用记录；
- 支持上链数据隐私安全保护。

6.8.2 应用场景

应用场景描述如下：

- 代理节点接收环境监测设备编号，到区块链证书管理中心进行证书申请，申请确认后，设备被标识为边缘节点（区块链轻节点）；
- 边缘节点对设备参数和设备监测值进行数字签名并主动或被动传输给区块链节点；
- 区块链节点对收到的边缘节点数据进行合法性校验，通过校验后附加认证信息上传到数链系统，完成监测设备数据上链。

6.8.3 应用场景图

区块链数据服务应用场景见图 11，生态环境智能质控物联网设备接入示例见附录 A。

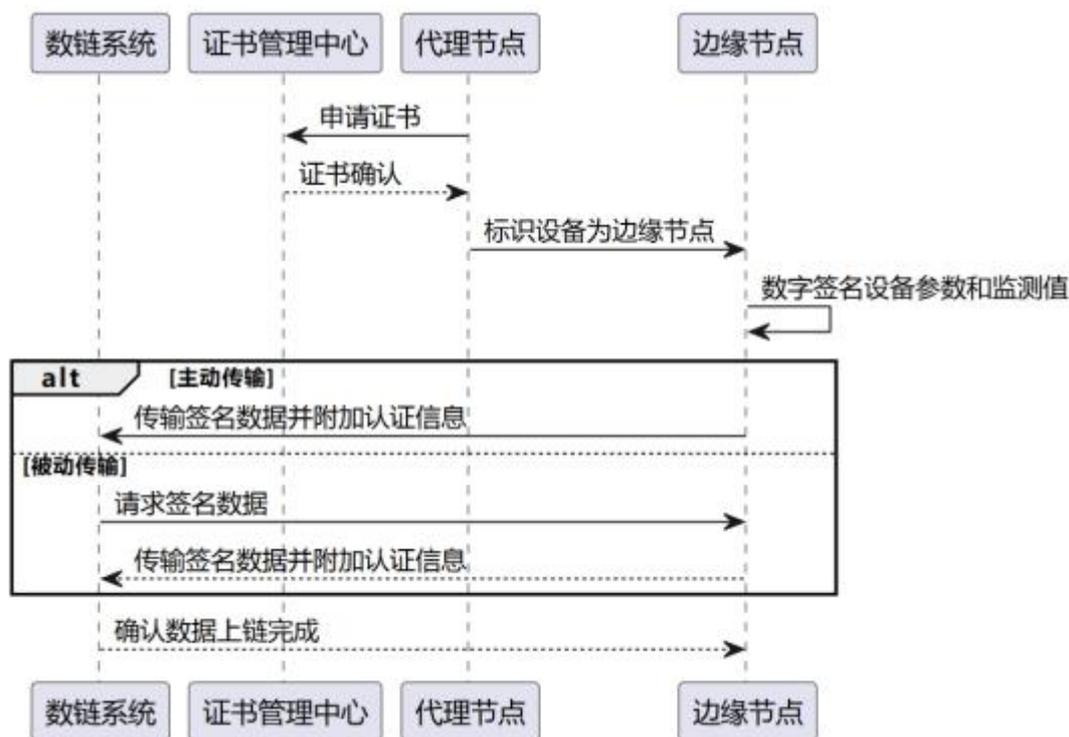


图 11 区块链数据服务应用场景

7 接入质量管理

本文件中的数据质量评价维度、数据接入质量技术要求和数据质量评分方法应符合DB15/T 1873-2020第5、6、7章的要求。

8 接入安全管理

8.1 数据源认证

外部公共数据源认证参照DB15/T 1874-2020第7.2章节进行数据源鉴别和记录，外部其他数据源遵从数据源身份认证要求。

内部数据源应采用区块链数字证书技术进行身份认证。

8.2 数据可靠传输

外部数据源遵从数据源要求的传输方式。

内部数据源之间进行数据传输时应采用区块链数字签名及时间戳方法保证数据传输的完整性、私密性、可溯性、不可抵赖性和时间证明。

附录 A (资料性) 智能质控生态环境物联网设备接入说明

A.1 申请数字证书

智能质控生态环境物联网设备数据接入流程如下：

- a) 在区块链系统证书管理页面上传设备通用唯一标识码、设备操作系统数据指纹、设备应用软件于系统参数数据指纹，提交数字证书申请；
- b) 区块链系统生成数字证书；
- c) 分配代理节点。

A.2 设备接入

设备接入流程如下：

- a) 设备接入区块链系统时需上传设备通用唯一标识码、设备操作系统数据指纹、设备应用软件与系统参数数据指纹；
- b) 区块链系统进行数据比对，比对一致方可允许设备接入；
- c) 下载数字证书；
- d) 下载代理节点参数；
- e) 上传本设备公钥；
- f) 上传本设备数据。

A.3 数据传输

示例：数据接口内容见表 A.1

表 A.1 数据接口内容示例

接口名称	数据上报
接口形式	https://{ip}:{port}/post/data
接口协议	WebService
接口方法	dataPost
接口描述	智能质控生态环境物联网设备监测数据上报

A.4 接口上传数据格式

示例：原始数据格式

```
{
  "deviceId": "XXX.....XXX",           //设备ID
  "devicePara": {
    "para1": "XXXXXX",                 //参数1
    "para2": "XXXXXX",                 //参数2
    .....
    "paraN": "XXXXXX"                  }, //参数N
  "measureValue": {
    "data1": "XXXXXX"
    "data2": "XXXXXX"
    .....
    "dataN": "XXXXXX"                  }
}
```

示例：数字签名后数据格式

```
{
  "deviceId": "XXX.....XXX",           //设备ID
  "dataFP": "XXX.....XXX",             //数据指纹，64字节
  "digitalSignature": "XXX.....XXX"    //128字节
}
```

参 考 文 献

- [1]GB/T 35274-2017 信息安全技术 大数据服务安全能力要求[S].
 - [2]GB/T 36345-2018 信息技术 通用数据导入接口[S].
 - [3]HJ 718-2014 环境信息共享互联互通平台总体框架技术规范[S].
 - [4]HJ 719-2014 环境信息系统数据库访问接口规范[S].
 - [5]HJ 721-2014 环境数据集加工汇交流程[S].
 - [6]HJ 727-2014 环境信息交换技术规范[S].
 - [7]DB15/T 1873-2020 大数据平台 数据接入质量规范[S].
 - [8]DB31/T 1240.2-2020 平台接入技术要求[S].
 - [9]DB35/T 1893-2020 生态环境大数据管理平台接口规范[S].
 - [10]DB36/T 1501-2021 生态环境大数据接入规范[S].
-