

ICS 35.240.01
CCS L 70
备案号 92462-2024

LD

中华人民共和国劳动和劳动安全行业标准

LD/T 6005—2023

人力资源社会保障部省数据同步技术规范

Technical specifications for data synchronization between ministries and provinces of
human resources and social security

2023 - 11 - 24 发布

2023 - 12 - 01 实施

中华人民共和国人力资源和社会保障部 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 数据同步管理	2
6 数据同步类型	2
6.1 数据同步场景	2
6.2 结构化数据	2
6.3 非结构化数据	3
6.4 半结构化数据	3
7 数据同步流程	3
7.1 数据库表同步	3
7.2 数据文件同步	5
8 平台接入规范	6
8.1 数据库表同步	6
8.2 数据文件同步	6
9 传输安全规范	8
10 运维管理规范	8
10.1 总体要求	8
10.2 运维事件分类	8
10.3 日常运维措施	8
10.4 突发故障处置	9
参考文献	10

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。

本文件由人力资源社会保障部提出并归口。

本文件起草单位：人力资源社会保障部信息中心。

本文件主要起草人：蒋旭卉、李晨星、马艳婷、张诚、张加会、汪诚、许莹莹、赵劲、洪祥、王岩、耿建军、徐梓川、孙永香、黄东东、闫包三。

人力资源社会保障部省数据同步技术规范

1 范围

本文件规定了部省两级人力资源社会保障部门之间开展纵向数据同步的技术要求，包括数据同步管理、数据同步类型、数据同步流程、平台接入规范、传输安全规范和运维管理规范。

本文件适用于人力资源社会保障部信息系统与各省、自治区、直辖市及新疆生产建设兵团人力资源社会保障部门信息系统（简称信息系统）之间的数据库（表）同步和数据文件同步。

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。

GB/T 2260 中华人民共和国行政区划代码

GB/T 39786—2021 信息安全技术 信息系统密码应用基本要求

LD/T 02.2—2022 人力资源社会保障电子认证体系规范 第2部分：电子认证系统技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据同步 data synchronization

当数据源发生改变时，其他相关数据也随之变化使之保持一致。

3.2

数据同步平台 data synchronization platform

通过集成第三方工具软件产品，依托人力资源社会保障业务专网（国家电子政务外网），建立跨层级数据传输链路，支持部省两级人力资源社会保障信息系统之间开展双向数据库（表）同步（简称数据库表同步）和数据文件同步的基础支撑平台软件。

注：数据同步平台包括人力资源社会保障部业务协同平台（以下简称业务协同平台）等。

3.3

平台方 platform builder

根据人力资源社会保障跨层级数据共享要求，负责将源数据库（源文件）传输、装载到目标数据库（目标文件），发布、授权给数据使用方，支持部省数据同步的数据同步平台及其管理单位。

3.4

数据提供方 data provider

将源数据库（源文件）加载到目标数据库（目标文件），共享给其他信息系统使用的信息系统及其管理单位。

注：源数据库（源文件）是该信息系统的输出。

3.5

数据使用方 data user

接收和使用数据的信息系统及其管理单位。

注：目标数据库（目标文件）是该信息系统的输入。

4 缩略语

下列缩略语适用于本文件。

BLOB: 二进制大对象 (Binary Large Object)
CLOB: 字符大对象 (Character Large Object)
HTML: 超文本标记语言 (Hypertext Markup Language)
IPv6: 互联网协议第6版 (Internet Protocol Version 6)
JSON: JavaScript对象简谱, 是一种轻量级的数据交换格式 (JavaScript Object Notation)
SSL: 安全套接层 (Secure Socket Layer)
TCP/IP: 传输控制协议/网际协议 (Transmission Control Protocol/Internet Protocol)
UTF-8: 8比特万国码的可变长度字符编码 (8-bit Unicode Transformation Format)
XML: 可扩展标记语言 (EXtensible Markup Language)

5 数据同步管理

人力资源社会保障部省数据同步由平台方、数据提供方、数据使用方共同参与。各方应配备专职工作人员, 负责操作执行、监督审核和运维管理。具体职责分工要求如下。

- a) 平台方:
 - 负责数据同步平台第三方工具软件产品选型;
 - 制定数据同步平台在部署实施、日常使用、运维管理等方面的技术规范;
 - 为数据同步平台的部署实施、日常使用、运行监控提供技术支持;
 - 保障数据同步平台正常运行;
 - 负责目标数据库 (目标文件) 的装载、发布和授权;
 - 负责数据转让安全。
- b) 数据提供方:
 - 支持数据同步平台在本地的部署实施;
 - 负责源数据库 (源文件) 的创建、注册与发布操作;
 - 负责源数据库用户的创建、授权;
 - 负责审查、确认有关人员对源数据库 (源文件) 的全部操作;
 - 负责数据存储、共享安全。
- c) 数据使用方:
 - 支持数据同步平台在本地的部署实施;
 - 负责申请源数据库 (源文件) 使用权限、订阅目标数据库 (目标文件) 操作;
 - 负责目标数据库用户的创建、授权;
 - 负责审查、确认有关人员的权限申请和订阅操作;
 - 负责数据存储、使用安全。

6 数据同步类型

6.1 数据同步场景

人力资源社会保障部省数据同步可分为结构化数据同步、非结构化数据同步和半结构化数据同步, 按照数据同步的时效性, 可分为实时同步、非实时同步。

各种类型数据开展部省数据同步, 在数据初始化阶段, 宜采用离线的数据导出、导入方式, 通过数据使用方的质量校验后, 再进行增量数据追加, 直至数据同步进入实时状态; 在日常数据同步过程中, 宜采用异步的、非侵入性的技术路线 (如基于日志的变化数据捕获), 降低对数据提供方业务运行的影响。

数据同步平台支持日志、时间戳、标志位、触发器、全表拷贝、数据比对等主流数据同步技术路线, 以适应不同的业务场景、不同的同步策略。

6.2 结构化数据

结构化数据是指由二维表结构来逻辑表达和实现的数据，严格遵循数据格式与长度规范，主要通过关系型数据库进行存储和管理。

结构化数据开展部省数据同步，可采用如下方式：

- a) 实时同步：借助数据同步平台进行数据库表同步；
- b) 非实时同步：数据提供方将需要传输的数据集从数据库（表）中导出，生成源文件，借助数据同步平台进行数据文件同步。

6.3 非结构化数据

非结构化数据是指数据结构不规则或不完整、没有预定义的数据模型、不方便用数据库二维逻辑表来表现的数据，包括所有格式的办公文档、文本、图片、HTML文件、各类报表、图像、音频/视频信息等。

非结构化数据开展部省数据同步，可借助数据同步平台进行非实时数据文件同步。

6.4 半结构化数据

半结构化数据是指结构变化很大的结构化数据，由于结构变化很大不能建立与之对应的表，同时由于需要获取数据细节，也不宜将数据组织成一个文件并按照非结构化数据处理，如日志、XML文件、JSON文件等。

半结构化数据开展部省数据同步，可借助数据同步平台进行非实时数据文件同步。

对于在某些数据库（表）中以CLOB/BLOB方式保存的大字段，以及保存在分布式存储中的数据（如采用分布式对象存储的电子业务材料），可选择非实时数据文件同步，也可由数据提供方开发服务接口，借助业务协同平台实现跨层级数据共享。

7 数据同步流程

7.1 数据库表同步

7.1.1 同步原理

数据同步平台为支持数据库表同步，提供源端数据同步代理工具和目标端数据同步代理工具组件，以及监控管理服务。

数据库表同步时，数据同步平台的源端数据同步代理工具从源数据库中，按照“先进先出”原则抓取需要传输的变化数据，将数据封装成一个由特定消息头（Header）和消息体（Body）组成的、固定结构的加密数据包，并传递给目标端数据同步代理工具；目标端数据同步代理工具在接收到数据包后，进行传输合法性、完整性、不重复性检查，再严格按照数据操作的顺序，装载数据到目标数据库。数据库表同步原理见图1。

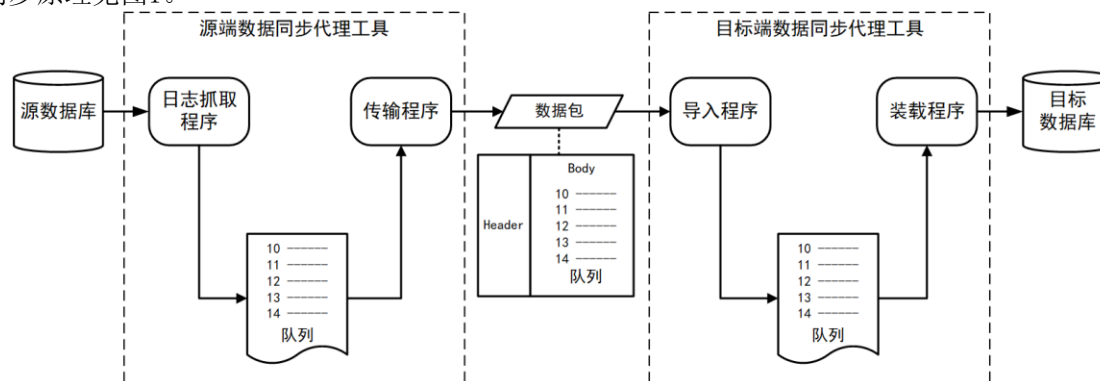


图1 数据库表同步原理

数据库表同步采用“行映射”的方式，实现变化数据的快速定位。具体方法如下：

- a) 源端数据同步代理工具从源数据库的日志文件中获取变化数据的操作信息，并把该操作记录的具体定位标识出来；
- b) 当变化数据在目标数据库装载时，通过保存在目标数据库中的行映射表，得到该数据的位置行信息，装载程序通过该信息能够直接定位变化数据的写入位置。

7.1.2 同步过程

7.1.2.1 概述

数据库表同步借助数据同步平台搭建部省数据传输链路，并在链路两端配置数据对象复制映射关系，实现源数据库和目标数据库的数据同步。

7.1.2.2 准备工作

开展数据库表同步，应完成如下准备工作：

- a) 资源准备：提供用于部署数据同步平台相关组件的服务器设备；
- b) 平台部署：在源数据库和目标数据库端分别部署数据同步代理工具，建立数据同步传输链路；
- c) 创建用户：在源数据库和目标数据库中，分别为数据同步代理工具创建数据库用户并授权；
- d) 数据结构同步：借助数据同步代理工具或人工执行数据库脚本方式，在源数据库和目标数据库之间同步数据库结构；
- e) 运行检测：在数据同步正式运行前，对数据库用户授权、数据库结构、数据同步传输链路进行测试。

7.1.2.3 实现过程

数据库表同步实现过程见图2。

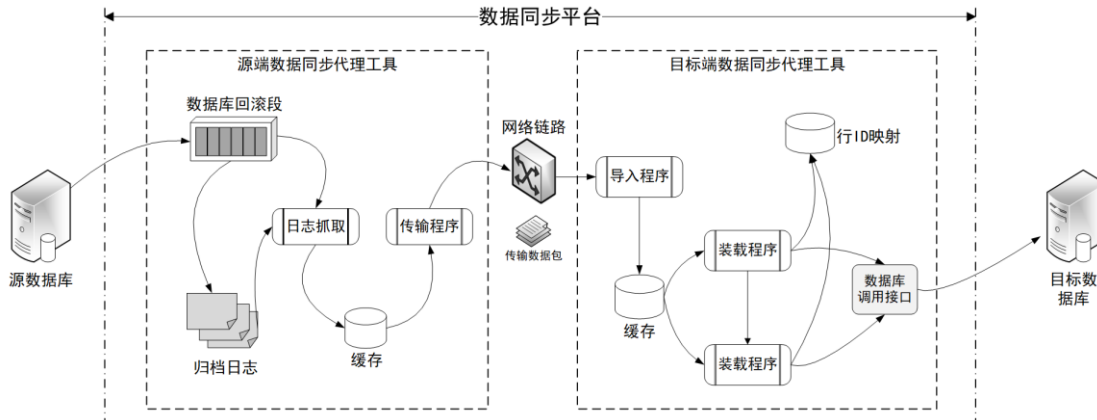


图2 数据库表同步实现过程

实现过程说明如下：

- a) 源数据同步：源端数据同步代理工具抓取源数据库日志；
- b) 操作日志分析：源端数据同步代理工具分析日志，获得同步数据的中间表示形式文件（联机日志缓存文件），包含已提交、未提交和回滚操作数据；
- c) 操作数据合成：对联机日志缓存文件按照操作时序、提交逻辑进行合成，舍弃回滚和未提交的操作数据，保留已提交的操作数据；
- d) 操作数据传输：操作数据先保存到缓存中，源端数据同步代理工具将其封装成数据包，通过TCP/IP协议传输至目标端数据同步代理工具，传输通道应通过SSL证书加密；
- e) 操作数据装载：目标端数据同步代理工具将数据包放入队列缓存，再依次装载操作数据；
- f) 目标数据装载：包括初始化数据装载和增量数据装载，目标端数据同步代理工具调用目标数据库接口，按照操作数据队列，写入同步数据；

- g) 数据校验：数据同步平台支持表级校验（源数据库和目标数据库中的表数据量是否一致）和字段级校验（源数据库和目标数据库中表的每个字段内容是否一致），保障数据库表同步的一致性。

7.2 数据文件同步

7.2.1 同步原理

数据同步平台为支持数据文件同步，提供源端服务节点、目标端服务节点和客户端辅助程序组件，以及监控管理服务。

数据文件同步时，数据同步平台通过源端服务节点、目标端服务节点提供目录及文件的收发服务。数据提供方指定发送目录和源文件，数据使用方确定接收目录，源端服务节点启动发送指令，将发送目录中的源文件（含子目录）以相同的目录结构发送到接收目录下，目标端服务节点进行传输合法性、完整性、不重复性检查，数据使用方通过接收目录获取目标文件，文件收发状态通过消息服务在服务节点间共享。数据文件同步原理见图3。

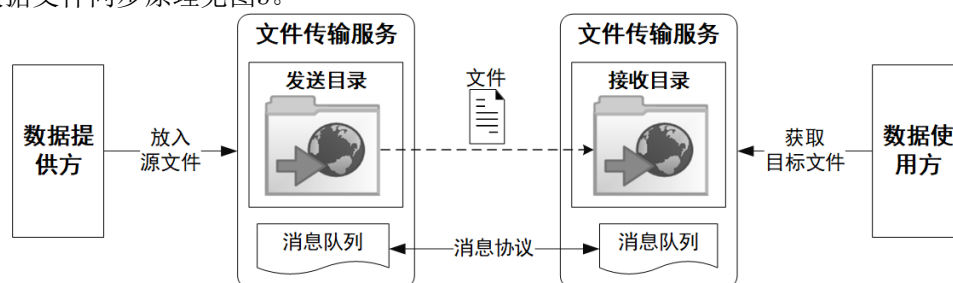


图3 数据文件同步原理

7.2.2 同步过程

7.2.2.1 概述

数据文件同步借助数据同步平台搭建部省数据传输链路，通过在服务节点上创建文件发送目录及接收目录，配置传输参数、任务管理和定时清理策略，实现源文件和目标文件的数据同步。

7.2.2.2 准备工作

开展数据文件同步，数据提供方和数据使用方在完成资源准备和平台部署后，分别在数据同步平台服务节点上创建文件发送目录和接收目录，定义任务，指定要发送的目录及文件、接收节点及接收目录，并对任务启停、执行方式、任务规则策略等进行配置。

7.2.2.3 实现过程

数据文件同步实现过程见图4。

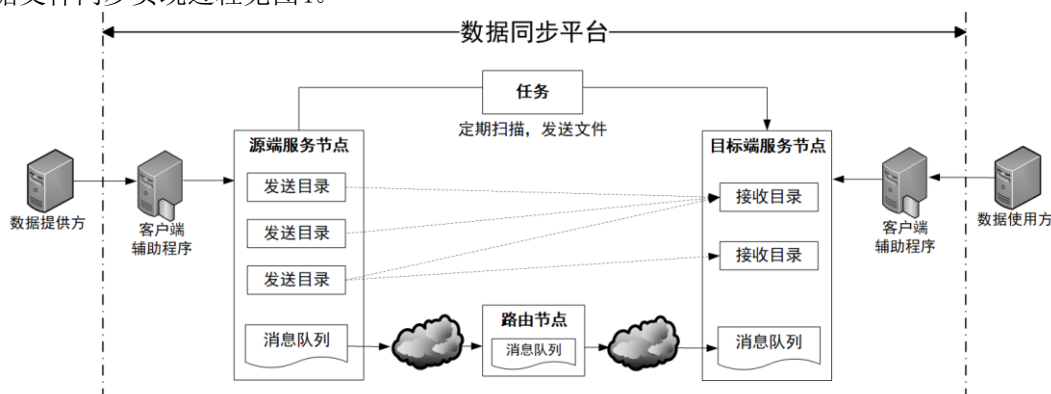


图4 数据文件同步实现过程

实现过程说明如下：

- a) 横向传输：数据提供方调用客户端辅助程序，执行文件传输命令，将源文件提交到源端服务节点发送目录；
- b) 纵向传输：源端服务节点启动任务，将源文件发送到目标端服务节点（支持压缩、加密），同时通过消息队列，同步任务状态到目标端服务节点，传输通道应通过 SSL 证书加密；
- c) 获取文件：数据使用方通过目标端服务节点获取目标文件发送状态消息，调用客户端辅助程序，从接收目录下获取目标文件（支持解压、解密、去重）。

8 平台接入规范

8.1 数据库表同步

8.1.1 基本要求

数据提供方和数据使用方接入数据同步平台，应遵循的基本要求包括：

- a) 提供数据同步代理工具部署运行所必要的软硬件基础设施和网络环境；
- b) 确保源数据库和目标数据库在数据库类型、数据类型、数据库服务器等方面满足平台技术要求；
- c) 确保数据同步链路畅通，网络传输速率达到平台基本要求并保持稳定；
- d) 在源数据库和目标数据库中，分别为数据同步代理工具创建数据同步用户，并进行授权；
- e) 源数据库和目标数据库的表结构应满足一致性要求，双方字符集编码不同时，调整目标端的字段长度，兼容超长字段；
- f) 在只进行数据同步时（不用于数据库迁移、数据库灾备等目的），应过滤掉源数据库的数据操纵语言(DML, Data Manipulation Language)和数据控制语言(DCL, Data Control Language)。

8.1.2 同步策略

数据库表同步支持全量同步和增量同步两种策略：

- a) 数据初始化同步，采用全量同步策略；
- b) 已经完成数据初始化目标数据库，采用增量同步策略。

8.1.3 同步频率

全量同步按照数据使用方的需要执行，增量同步根据库表数据变化情况实时进行。

8.1.4 校验机制

全量同步完成后、增量同步过程中，支持通过表级校验、字段级校验来验证数据一致性。校验频次和校验策略可根据业务需要酌情配置。

8.1.5 异常处理

数据同步平台通过监控管理服务开展全链路异常追踪，对数据同步代理工具异常、网络异常、源数据库异常、目标数据库异常等进行记录和告警，支持断点续传，预留消息服务接口用于通知相关单位。

数据提供方宜启用源数据库“最小粒度”的日志记录模式，全面记录源数据库的各类操作，并妥善保管数据库归档日志，以便在异常状况修复后，平稳过渡到实时同步状态。

8.2 数据文件同步

8.2.1 基本要求

数据提供方和数据使用方接入数据同步平台，开展数据文件同步，应遵循的基本要求包括：

- a) 提供服务节点部署运行所必要的软硬件基础设施和网络环境；
- b) 数据提供方和数据使用方集成了平台客户端辅助程序；
- c) 确保数据文件的大小、类型、格式满足平台技术要求；
- d) 确保数据同步链路畅通，网络传输速率达到平台基本要求并保持稳定。

8.2.2 同步策略

数据文件同步支持全量同步和增量同步两种策略，可选择三种同步模式：

- a) 初始启动模式：随数据同步平台文件传输服务启动而启动，中途异常退出，监控管理服务也会将其重启；
- b) 定时启动模式：在指定的时间点启动文件传输，包括每年某时、每月某时、每天某时和特定四种模式；
- c) 按需启动模式：按照需要随时启动文件传输。

8.2.3 同步频率

全量同步按照数据使用方的需要执行，采用按需启动模式；增量同步采取定时启动或按需启动模式，定时任务宜安排在业务闲暇时（如晚23:00时后）启动。

8.2.4 命名规范

数据文件同步时，文件名、文件目录、传输任务遵循统一的命名规范。具体规范如下。

- a) 上传文件命名：采用“行政区划代码+应用系统名称+时间.文件后缀”（见示例1）。其中：
 - 行政区划代码遵循GB/T 2260；
 - 应用系统名称采用系统全称的汉语拼音首字母；
 - 时间采用系统时间（yyyyMMdHHmmss）。
- b) 分割文件命名：采用“行政区划代码+应用系统名称+时间+三位流水号.文件后缀”，三位流水号从001至999依次生成（见示例2）。
- c) 文件格式：交换文件统一采用UTF-8编码，其它编码格式文件应预先转换成UTF-8编码，再参与文件交换。
- d) 传输目录命名：文件传输目录应输入绝对路径（从根目录起开始输入，见示例3），目录名中不使用双引号（"”）、尖括号（<>）和竖线（|）等特殊字符，总目录名（包括文件名）长度宜控制在256字节（Byte）以内。
- e) 文件服务节点目录结构：数据同步平台的文件服务节点为每一个数据提供方建立独立的发送目录和接收目录，目录结构宜统一、对称（见示例4）。
- f) 文件传输任务命名：每一个文件传输任务都应独立命名，可使用中文、英文字母、数字字符、下划线组合命名（见示例5），最大长度不宜超过48字节（Byte）。

示例1：

山东省联网监测系统提交的某份源文件命名为：
370000LWJCXT20211212162030.doc。

示例2：

山东省联网监测系统提交的某份源文件分割后生成的第13个子文件命名为：
370000LWJCXT20211212162030013.doc。

示例3：

业务协同平台的根目录输入规范：
——Linux/Unix下，根目录为：/home/yzzx；
——Windows下，根目录为：driver\yzzx，driver可以为“C:”、“D:”等。

示例4：

业务协同平台在部署实施过程中，使用数据库/系统简称的汉语拼音首字母作为数据提供方的标识，并分别建立发送目录和接收目录。如联网监测系统（使用“lwjc”标识）的目录结构如下：
——发送目录：/lwjc/send/；
——接收目录：/lwjc/recv/。

示例5：

业务协同平台的文件传输任务分为省级系统上报任务和部级系统下发任务。任务命名规则如下：

- 省级系统上报任务：省份/省级单位简称_数据库/系统简称数据上报；
- 部级系统下发任务：部_数据库/系统简称数据下发_省份/省级单位简称。

如新疆生产建设兵团联网监测系统、人力资源社会保障部联网监测系统之间的双向文件传输任务命名如下：

- 上报任务：兵团_联网监测数据上报；
- 下发任务：部_联网监测数据下发_兵团。

8.2.5 异常处理

数据同步平台通过监控管理服务开展全链路异常追踪，对客户端辅助程序异常、网络异常、源端服务节点异常、目标端服务节点异常等进行记录和告警，支持断点续传，支持启动补偿机制，预留消息服务接口用于通知相关单位。

8.2.6 校验机制

数据同步平台提供批量数据文件同步对账服务，支持采用汇总对账和明细对账相结合方式，来验证数据文件一致性。对账策略可根据业务需要酌情配置。

8.2.7 补偿机制

数据文件同步传输任务失败后，数据同步平台提供至少3次重发服务。若一个文件（含子目录）重发服务完成后仍未成功，则任务失败，放入失败目录。

数据同步平台监控管理服务支持查看成功任务、失败任务，手动选择失败任务重新传输。

9 传输安全规范

数据同步平台提供多种策略，保障数据传输过程的安全、可靠。主要措施包括：

- a) 密钥管理：可按照 GB/T 39786—2021、LD/T 02.2—2022 的规定，借助人力资源社会保障电子认证体系，统一生成、分发、管理加密密钥；
- b) 传输加密：支持传输数据（文件）加密、传输通道加密，支持国家密码主管部门认可使用的密码算法（国产密码算法优先），支持接入第三方安全策略；
- c) 传输日志：传输过程中记录完整的流水日志；
- d) 网络协议：支持 IPv6 协议，满足 IPv6 规模部署应用要求；
- e) 限速限流：提供传输流量限制功能，避免争用带宽。

10 运维管理规范

10.1 总体要求

平台方、数据提供方、数据使用方按照“谁部署、谁维护、谁负责”的原则，完善本地运维规范，提高部省数据同步的应急处置能力，防止因系统故障造成的数据丢失、泄露，保障数据同步平台的正常运行。

10.2 运维事件分类

运维事件分日常运维和突发故障两类。

- a) 日常运维：保障数据同步平台正常运行的日常维护工作，包括接入信息系统审核与联调，数据同步平台程序故障（BUG）修改、版本更新、定期巡检等。
- b) 突发故障：突发的、影响数据同步平台正常运行的故障，包括服务器硬件故障、通讯线路设备故障、通讯线路故障、数据库故障、软件系统故障和异地通讯故障等。

10.3 日常运维措施

数据同步平台运行过程中，平台方、数据提供方、数据使用方应开展的工作包括：

- a) 接入新的信息系统时，审慎评估系统的技术合规性；
- b) 数据同步平台出现程序故障（BUG）时，及时分析、修复；
- c) 数据同步平台版本更新时，及时安装部署更新，同步做好程序备份，支持更新回滚；
- d) 定期开展数据同步平台、基础设施环境、网络通道的巡查预警。

10.4 突发故障处置

当突发故障时，平台方、数据提供方、数据使用方应开展的处置措施包括：

- a) 本地服务器硬件、通讯线路设备硬件问题，及时启用备份系统；
- b) 本地通信线路问题，及时与网络管理员和运营商联系并启用备份线路；
- c) 本地数据库的问题，及时启用备份数据库，同时联系数据库供应商及时解决；
- d) 异地通讯故障，联系故障地的平台运维保障人员，及时查找故障点并寻求解决方案。

参 考 文 献

- [1] GB/T 28827.1—2012 信息技术服务 运行维护 第1部分：通用要求
 - [2] GB/T 35273—2020 信息安全技术 个人信息安全规范
 - [3] 人力资源社会保障行业信息化领域密码应用实施方案（2020-2022年）
-