

ICS 点击此处添加 ICS 号
点击此处添加中国标准文献
分类号

T/CEC

中国电力企业联合会团体标准

T/CEC XXXX—XXXX

电力管理信息系统实时数据采集技术规范

Technical specification for real-time data acquisition of power management
information system

2024—XX—XX 发布

2024—XX—XX 实施

中国电力企业联合会发布

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	2
5 总则	3
6 结构化数据采集方案	3
7 非结构化数据采集方案	5
8 时序数据采集方案	8
9 实时计算框架技术架构	12
10 系统保障及应用	13

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》给出的规定起草。

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。

本文件由中国电力企业联合会提出。

本文件由中国电力企业联合会知识管理标准化技术委员会（CEC/TC 33）归口。

本文件起草单位：

本文件主要起草人：

本文件首次发布。

本文件在执行过程中的意见或建议反馈至中国电力企业联合会标准化管理中心（北京市白广路二条一号，100761）

电力管理信息系统实时数据采集技术规范

1 范围

本文件规定了不同类型数据的采集技术路线、采集方式、采集实施工作及实时计算框架应用等技术要求，不涉及网络相关标准及要求。

本文件适用于电力管理信息系统（包括企业级管理信息系统、计量系统、调度系统、生产域系统、专业子公司专业系统等）的结构化、非结构化数据和时序数据的实时数据采集工作。

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 8567 计算机软件文档编制规范

GB/T 35295-2017 信息技术 大数据 术语

GB/T 9385-2008 计算机软件需求规格说明规范

GB/T 9386-2008 计算机软件测试文档编制规范

GB/T 11457-2006 信息技术 软件工程术语

GB/T 13730-2002 地区电网调度自动化系统

GB/T 14394-2008 计算机软件可靠性和可维护性管理

GB/T 15532-2008 计算机软件测试规范

GB/T 22032-2021 系统与软件工程 系统生存周期过程

GB/T 36073—2018 数据管理能力成熟度评估模型

GB/T 38619-2020 工业物联网 数据采集结构化描述规范

GB/T 39788-2021 系统与软件工程 性能测试方法

T/GZBD 5-2020政务数据采集规范

ISO/IEC 27001:2013 信息安全管理体系

ISO/IEC 27034-7 2018 软件安全标准

NHJC -10.01-2018 重点用能单位能耗在线监测数据采集指南 第 10-1部分 电力行业数据采集指南

3 术语和定义

GB/T 35295-2017和GB/T 11457-2006界定的以及下列术语和定义适用于本文件。

3.1

电力数据 power system data

通过传感器、智能设备、视频监控设备、音频通信设备、移动终端等各种数据采集渠道收集到的海量数据集合，这些数据来源于发电、输电、变电、配电、用电和调度等各环节，涵盖了结构化、半结构化、非结构化等多种类型。既包括发电数据、电网数据、用电数据、能源数据等电力企业自身运营的相关数据，也包括天气、市场需求、社会环境等相关的外部数据。

3.2

数据中心 data center

由数据中心向各业务应用提供统一的数据库环境，业务应用的TP型和AP型数据存储和访问，均遵循数据中心统一标准，在数据中心提供的数据库内统一完成，业务系统本地不再存储TP型数据。数据中心负责业务数据区（TP）向分析数据区（AP）的数据同步，确保TP型数据和AP型数据的准确、一致、实时联动，既支持分析型应用（AP），也支持事务型应用（TP）。

3.3

实时数据 real-time data

通过传感器等物联网技术实时采集、传输的，用于描述业务过程和设备状态当前或非常短时间之内的数据。它可以有不同的存储地点，比如移动设备、系统内存等，它的数据状态是更新的、及时的，以及可以提供实时反馈服务的。这些数据具有时效性强、更新速度快的特点，对于实时监控、预测预警和决策支持等应用具有重要意义。

3.4

实时计算 real-time calculation

针对实时产生的数据进行立即处理和分析的计算方式。这种计算方式要求系统能够在数据产生的同时，快速、准确地完成计算任务，并返回结果。实时计算通常涉及对海量数据的处理，并且要求处理速度达到秒级甚至毫秒级，以满足对实时性的高度需求。常用的实时计算框架有Apache Flink、实时计算框架 Streaming和Apache Storm等。

3.5

分布式文件系统 distributed file system

多个结构化数据集分布在一个或多个服务器集群的各个计算机节点的文件系统。

注：此类系统中，数据可能分布在文件和/或数据集层，更为普遍的是在数据块这个层级分布，同时支持集群中多个节点和大型文件和/或数据集的不同部分交互。

[来源：GB/T 35295-2017，2.1.21]

3.6

容器集群管理系统 container cluster management system

实现容器集群的自动化部署、自动扩缩容、维护等功能。

3.7

任务 task

由一个线程执行处理任务。

3.8

电力管理信息系统 power management information system

专门为电力企业设计的集成化信息系统，旨在通过信息技术手段，实现对电力生产、传输、分配和消费等各个环节的全面、高效管理。系统以计算机硬件、软件和网络通信设备为基础，通过收集、整理、分析电力数据，提供决策支持、优化资源配置、提升运营效率，以及保障电力系统的安全稳定运行。其范围涵盖电力生产、传输、分配和消费等全过程，包括电力生产、设备资产、市场营销、能源管理、安全管理、决策支持等方面的系统。

4 缩略语

JVM: Java虚拟机 (Java Virtual Machine)

GC: 即垃圾回收 (Garbage Collection)

SQL: 结构化查询语言 (Structured Query Language)

SCADA: 数据采集与监控系统 (Supervisory Control And Data acquisition)

API：应用程序接口（Application Programming Interface）

CDC：变化数据获取（Change Data Capture）

5 总体要求

5.1 实时性

对通过系统传输的每个数据项执行操作，实时统计获取计算结果。

5.2 高性能

采用高吞吐、高性能、低时延的实时流处理引擎，分布式集群部署，能够提供毫秒级时延处理能力。

5.3 高可用性

集群出现机器故障时，结合分布式程序协调服务器对故障机器进行恢复，能够重新协调任务执行，通过分布式存储引擎提供数据恢复，最终保证任务的恢复。

5.4 适用业务系统

适用本文件的业务系统及数据类型见表1。

表 1 业务系统范围

业务域	业务系统	结构化数据	非结构化数据	时序类数据
管理域	营销类系统	√	√	
	生产管理类系统	√	√	
	财务类系统	√	√	
	人力资源管理类系统	√	√	
	协同办公类系统	√	√	
	培训与评价类系统		√	
	预警监测类系统		√	
	IT 呼叫类系统		√	
计量域	计量自动化类系统	√		√
调度域	调度自动化类系统	√		√
	配网自动化类系统	√		√
生产域	监控视频类系统		√	
	机巡类系统		√	
	SCADA			√
新兴业务域	电力交易类平台	√	√	
	电子商城类系统	√	√	
	融资租赁类系统	√		
	金融类系统	√		
	水工水情类系统	√	√	
	发电设备可靠性类系统	√		
	储能类系统	√	√	

6 结构化数据采集

6.1 技术路线

结构化数据可根据业务系统数据量、业务并发量、负载情况等因素，选择是否通过备份库/管理库采集。通过CDC软件将源端的变更数据采集到数据中心的CDC解析服务器，再推送至消息队列，通过实时消费数据，将数据写入数据中心。

互联网区数据实时采集和内网数据实时采集的区别主要是互联网区数据需要把数据同步到DMZ区服务器，再通过内网交换平台，把数据从DMZ区同步到内网，再发送到消息队列，最后通过流计算将数据写入到数据中心。

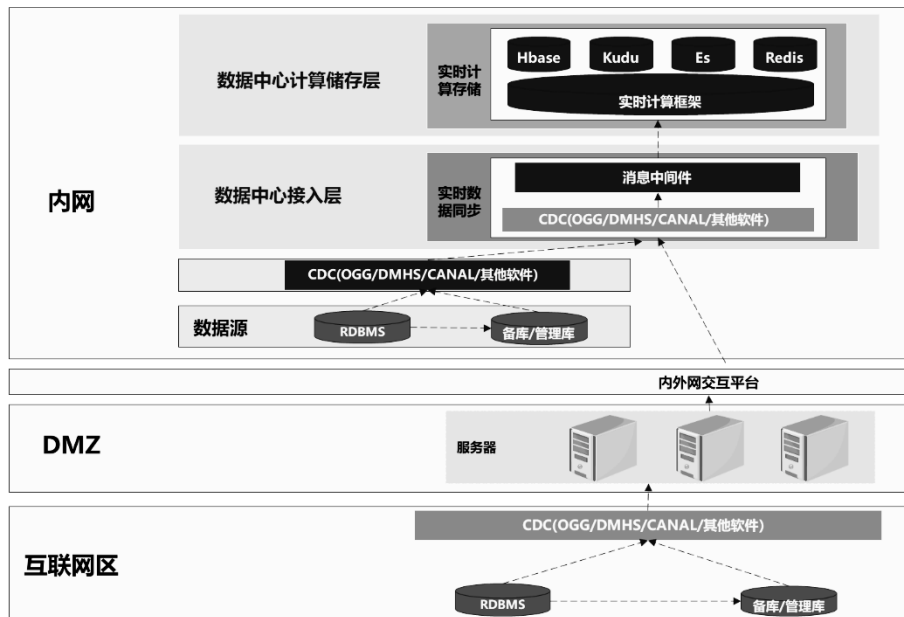


图 1 结构化数据采集技术路线图

6.2 采集方式

根据业务系统数据量、业务并发量、负载情况，各业务系统结构化数据采集方式见表 2。

a) 读写分离模式：业务系统侧为业务系统源库做读写分离，建立业务库的只读库/管理库，并使用CDC进行数据同步。在只读库/管理库上再部署CDC同步软件，并配置任务将解析后的日志文件投递至数据中心采集服务器；

b) 直连模式：CDC同步软件部署在业务系统生产库上，配置任务将文件实时投递至数据中心解析服务器。

表 2 各业务系统结构化采集方式

业务域	业务系统	采集方式
管理域	营销系统	读写分离模式
	生产管理系统（生产、投资计划、基建、项目管理、物资、电子商务）	读写分离模式
	财务系统	直连模式
	人力资源管理系统	直连模式
	协同办公系统	直连模式

计量域	计量自动化系统	读写分离模式
调度域	调度自动化系统	读写分离模式
	配网自动化系统	读写分离模式
新兴业务域	电动汽车充电服务平台	读写分离模式
	电力交易平台	读写分离模式
	电子商城系统	读写分离模式
	融资租赁系统	读写分离模式
	金融业务系统	读写分离模式
	水工水情类系统	读写分离模式
	发电设备可靠性类系统	读写分离模式
	储能类系统	读写分离模式

6.3 采集要求

电力管理系统中结构化数据采集要求如下：

数据采集频率：指数据采集的时间间隔，通常以秒为单位。对于电力管理系统，数据采集频率需要根据实际需求进行选择，以保证数据的实时性和准确性。

数据传输时延：指从数据采集到数据传输到数据中心或其他目的端的时间。数据传输时延越短，数据的实时性越好，对于电网行业实时数据采集应达到60秒。

数据一致性：指采集到的数据的准确程度，包含数据的内容、数据精度的准确程度。数据准确度越高，数据的可靠性越好。电网行业实时数据采集应达到98%以上。

数据完整性：指采集到的数据是否完整，包含数据的条数、字段没有遗漏或缺失的情况。数据完整性是数据处理的必要条件。电网行业实时数据采集应达到98%以上。

数据安全性：指数据采集系统能够保证数据不被非法获取或篡改的能力。数据安全性是数据采集系统的重要指标之一。数据采集系统应符合系统安全等保三级的防护要求。

系统稳定性：指数据采集系统在运行过程中是否稳定可靠，能否保证持续的正常运行。系统稳定性是数据采集系统的重要指标之一。电网行业实时数据采集系统年度系统可用性应在98%以上。

7 非结构化数据采集

7.1 技术路线

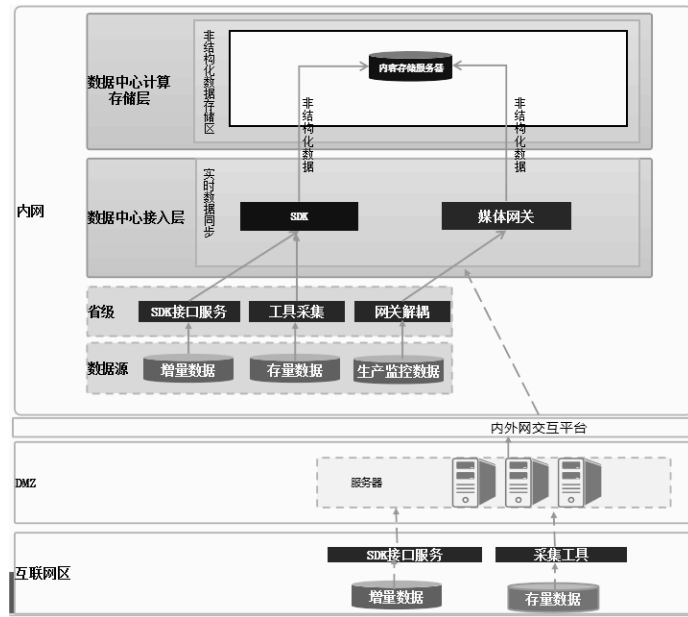


图 2 非结构化数据采集技术路线图

非结构化数据可根据业务系统的存量数据、增量数据、生产监控视频数据来选择通过SDK接口服务、工具采集、网关解耦等方式进行采集，通过SDK接口和媒体网关将源端的数据采集到数据中心。

新建系统及平台应采用SDK开发方式接入，并在设计、初验前实施数据管控，存量业务系统推荐按SDK开发方式接入，若现实不具备条件则采用工具采集。

7.2 采集方式

7.2.1 通用方式

各类业务系统接入到非结构化数据服务平台，提供“SDK开发接入”、“工具采集”、“网关解耦”三种接入方式。

- a) SDK开发接入：改造源端系统，主要针对增量非结构化数据的实时接入；
- b) 工具采集：主要针对历史存量非结构化数据批量接入；
- c) 网关解耦：主要针对生产监控视频流数据的接入。

表 3 各系统采集方式

业务域	业务系统	采集方式
管理域	营销系统	工具采集
	生产管理类系统	工具采集
	财务系统	工具采集
	人力资源管理系统	工具采集
	培训与评价系统	工具采集
	招聘系统	工具采集
	协同办公系统	工具采集
	网络信息预警监测系统	工具采集
	呼叫系统	工具采集

生产域	监控视频系统	网关解耦
	机巡系统	工具采集
	物联网	SDK
新兴业务域	电力交易平台	SDK
	电子商城系统	SDK

7.2.2 SDK 开发接入

“SDK开发接入”方式是指业务系统通过调用非结构化数据服务平台提供的SDK实现实时接入，须按非结构化数据服务平台的SDK开发规范对业务系统所有涉及非结构化数据的存储、访问、检索等功能进行代码改造，以HTTP的方式将非结构化数据传输到平台存储。

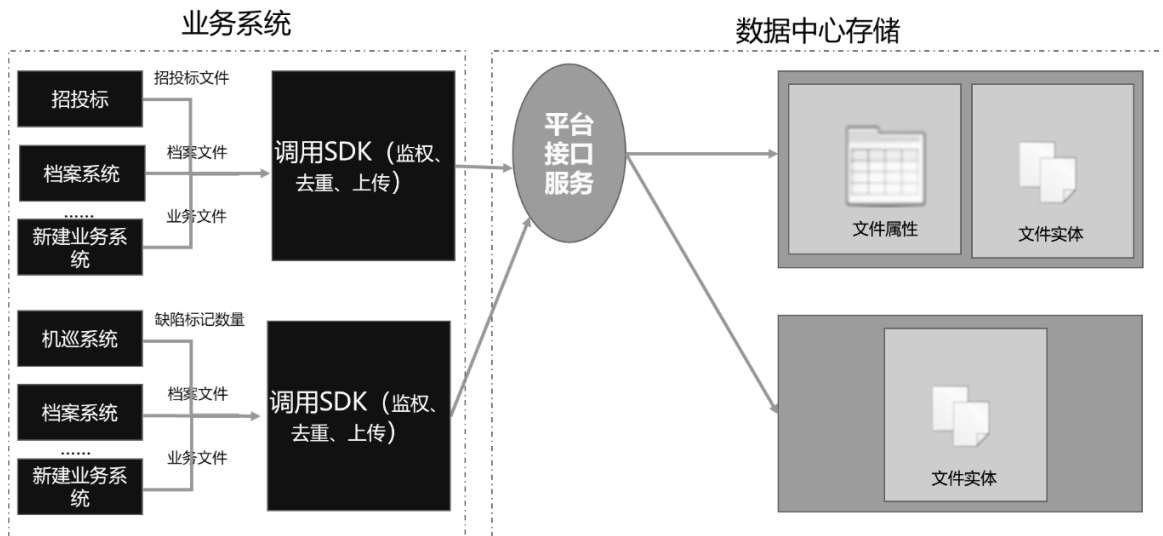


图 3 非结构化数据“开发接入”方式

非结构化数据“开发接入”方式如图3所示：

- 1) 业务侧系统调用非结构化数据服务平台提供的SDK，将非结构化数据的文件实体以及业务属性的存储至非结构化数据服务平台；
- 2) 非结构化数据服务平台返回唯一码，后续业务系统可使用该唯一码获取文件；

7.2.3 工具采集接入

“工具采集”方式是指通过在业务侧文件服务器上安装非结构化数据采集工具完成非结构化历史存量数据的采集工作。

非结构化数据“工具采集”方式如图4所示：

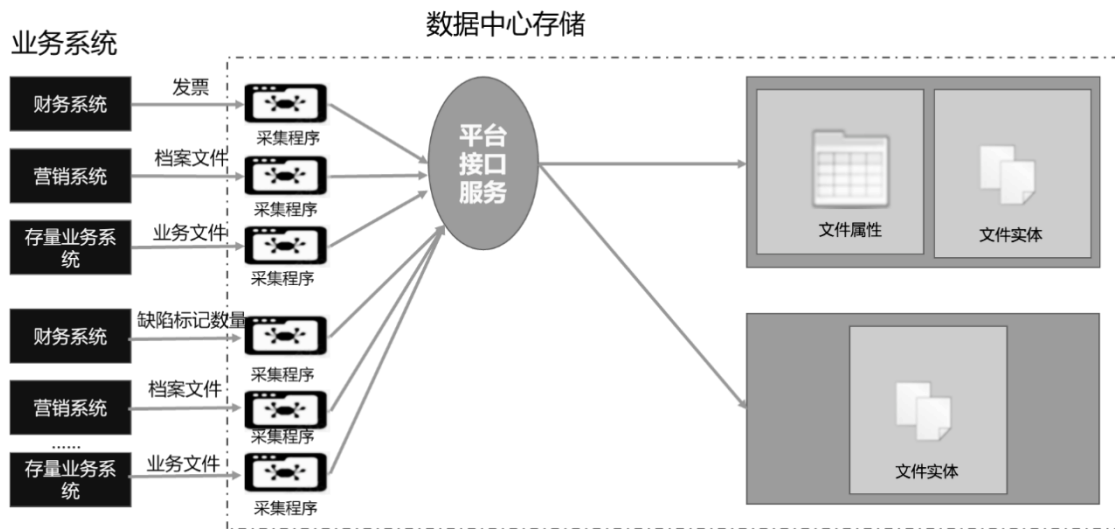


图 4 非结构化数据“采集工具”方式

1) 业务侧根据“SDK开发接入”确定的非结构化数据的业务属性梳理业务视图，确保存量数据与增量数据的业务属性值一致；

2) 非结构化数据服务平台通过采集工具将业务侧系统存量非结构化数据的文件实体与业务属性采集至非结构化数据服务平台。

7.2.4 网关解耦方式接入

非结构化数据服务平台主要对接各级的视频主站，采用“网关解耦”方式进行接入，满足在非结构化数据服务平台进行监控视频数据实时播放和历史回放需求。其中主站需兼容PGCP、PGSP、RTSP、RTMP协议或GB/T 28181标准协议交互。

非结构化数据服务平台通过适配标准协议，封装API接口来提供视频的实时播放、视频录像回放等对外的业务能力。

7.3 采集标准

数据采集频率：指数据采集的时间间隔，通常以秒为单位。对于电力管理业务系统，数据采集频率需要根据业务附件大小、网络带宽进行评估，以保证数据的实时性和准确性。

数据传输时延：指从数据采集到数据传输到数据中心或其他目的端的时间。数据传输时延越短，数据的实时性越好，对于电网行业实时数据不超于100M的附件采集应达到60秒以内。

数据一致性：指采集到的数据的准确程度，包含数据的内容、附件大小的准确程度。数据准确度越高，数据的可靠性越好。电网行业实时数据采集应达到98%以上。

数据完整性：指采集到的数据是否完整，包含数据的条数、附件没有遗漏或缺失的情况。数据完整性是数据处理的必要条件。电网行业实时数据采集应达到98%以上。

数据安全性：指数据采集系统能够保证数据不被非法获取或篡改的能力。数据安全性是数据采集系统的重要指标之一。数据采集系统应符合系统安全等保三级的防护要求。

系统稳定性：指数据采集系统在运行过程中是否稳定可靠，能否保证持续的正常运行。系统稳定性是数据采集系统的重要指标之一。电网行业实时数据采集系统年度系统可用性应在98%以上。

8 时序数据采集方案

8.1 技术路线

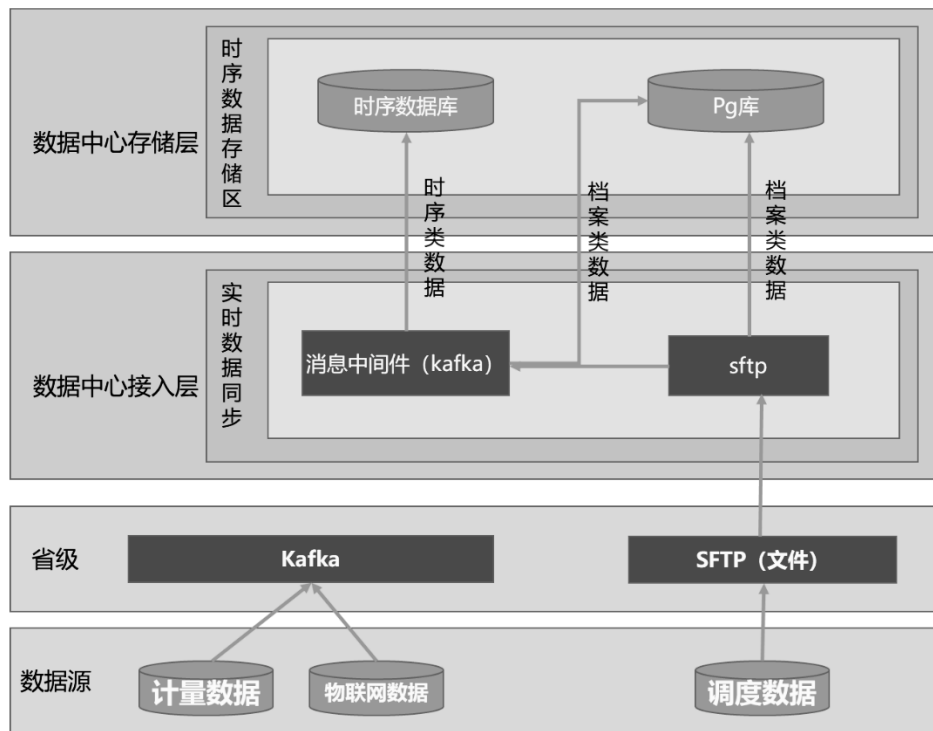


图 5 时序数据采集技术路线图

时序数据采集是依据业务系统数据量、业务系统相关要求和数据时效性来确定数据的采集方式。针对跨区数据传输，受限于安全II区与安全III区之间正反向物联隔离装置的限制，II区到III区的跨区数据传输只支持单向SFTP传输协议，所以目前主要采用e文件的传输方式，业务系统将文件推送至数据中心侧，数据中心解析文件后再根据不同的数据类型写入对应的库表中；针对数据实时性高的数据类型，目前主要采用消息总线的传输模式，业务系统通过消息队列将数据推送至数据中心，数据中心通过配置入库程序再分别写入时序数据库和关系型数据库。

8.2 采集方式

8.2.1 通用方式

各业务系统在系统数据项上要满足大数据平台数据采集规范，在数据采集方式上采用文件接入方式和消息队列数据接入方式。

8.2.2 文件数据采集方式

文件数据接入方式是采用SFTP方式由业务系统端（源端）按照约定的数据文件格式生成数据文件，并将数据文件推送到实时数据平台制定的文件服务器和对应的文件目录下。通过文件解析程序和配置入库程序将数据写入时序数据库和关系型数据库。

8.2.2.1 E 文件规范

E 文件的文件格式规范如下：

- 1) 在 E 文件中，以只含有“<类型名称>”字符串的行表示每一类型实时数据的开始和结束；
- 2) @开头表示字段行，#开头表示数据行，用 TAB 表示数据分隔；
- 3) 文件编码格式采用 UTF-8。

8.2.2.2 CIM 文件规范

依据CIM模型规范标准，设备全量模型文件（XML文件）一定要确保包含E文件中的各类设备数据，并且标签类完整。

8.2.2.3 普通数据文件规范

普通数据文件是除调度系统E文件之外，采用文件方式传输数据的文件格式。

- 1) ||(双竖线)表示数据分隔符；
- 2) 文件编码格式采用UTF-8。

8.2.2.4 文件命名规范

一级公司名称中文拼音首大写字母+“_”+二级公司中文拼音首大写字母+“_”+来源系统大写英文缩写+“_”+断面时间（格式：YYYYMMDD_HH24MISS）+“.”+文件类型后缀。

例如AA公司BB分公司SCADA系统2015年3月13日 下午5点5分7秒生成的数据文件（E文件）的文件命名：AA_BB_SCADA_20150313_170507.DT。

8.2.2.5 文件推送频率

依据数据采集规范要求，文件推送频率逐步实现数据实时推送，档案数据按照一周一次推送。

8.2.2.6 安全要求规范

应满足以下主要要求：

- 1) 按照业务域开通对应用户账号对文件读取权限控制。按照数据管控流程进行数据权限申请；
- 2) 不越权处理其他业务域的文件；
- 3) 文件类型必须遵从约定的格式；
- 4) 文件上传量必须遵从约定的数量，不能恶意上传垃圾文件；
- 5) 端口和策略开通需要明确，不开通无关的协议和端口。

8.2.2.7 运行保障

1) 非功能要求

SFTP数据文件复制应满足以下非功能性要求：

a) 在千兆带宽的网络条件下，数据复制吞吐量不少于50MB/秒（如果有多层网络穿会影响传输速度）。

2) 应用要求

SFTP数据文件方式进行数据采集应用应该满足如下要求：

- a) 业务系统必须开通TCP协议的访问策略；
- b) SFTP目录的读写权限需要单独申请和审批；
- c) 不提供删除、修改、剪切权限，仅提供读取和拷贝权限；
- d) 业务系统服务器必须有足够的存储空间。

8.2.3 消息队列数据采集方式

消息队列数据接入方式适用于大数据中心从业务系统获取数据量大、传输频率高、实时性很强的数据。

8.2.3.1 数据采集方式

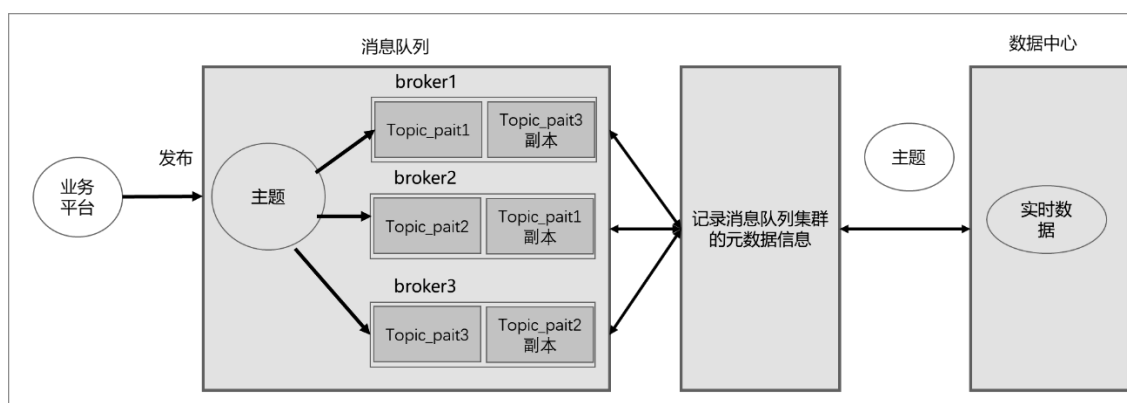


图 6 基于消息队列数据流的数据采集流程

业务系统通过消息队列客户端，将业务系统数据推送到规定的主题。时序数据库通过数据摄入组件从消息队列主题中获取数据后存储到相应的数据表中。

8.2.3.2 消息队列主题命名规则

消息队列的主题命名统一命名并对外发布，在接入过程中需要遵从如下规则：

系统名称+“_”+业务域+“_”+数据类型+“_”+一级公司代码+“_”+二级公司代码+“_”。

例如：TMR_CJ_YXNDL_03，TMR为系统名称，CJ为业务域（计量采集数据），YXNDL（运行电能量表数据），03（一级公司编码）。

8.2.3.3 运行保障

1) 消息队列数据流应满足以下非功能性要求：

a) 数据请求响应：不少于1万条/秒；数据抽取吞吐量：在千兆带宽的网络条件下，数据复制吞吐量不少于30MB/秒。

2) 消息队列数据流方式进行数据发布应用应该满足如下要求：

业务系统必须遵从消息队列的技术协议；

业务系统需要申请相应的策略和协议。

3) 安全要求

接口开通需要先申请，批准后接入；

按主题开通访问权限；

启用消息队列的安全认证机制；

端口和策略开通需要明确，不开通无关的协议和端口。

8.3 采集标准

数据采集频率：指数据采集的时间间隔，通常以秒为单位。目前电网运行时序数据的采集频率以与源端系统数据采集频率保持一致。

数据传输时延：指从数据采集到数据传输到数据中心或其他目的端的时间。数据传输时延越短，数据的实时性越好，对于电网行业实时数据采集应达到60秒。

数据一致性：指采集到的数据的准确程度，包含数据的内容、数据精度的准确程度。数据准确度越高，数据的可靠性越好。电网行业实时数据采集应达到98%以上。

数据完整性：指采集到的数据是否完整，包含数据的条数、字段没有遗漏或缺失的情况。数据完整性是数据处理的必要条件。电网行业实时数据采集应达到98%以上。

数据合规性：指采集到的数据是否符合业务规范，目前时序数据采集过程中会从数据层面发现一些不符合业务规则的数据。电网行业实时数据的合规率应达到99%以上。

数据安全性：指数据采集系统能够保证数据不被非法获取或篡改的能力。数据安全性是数据采集系统的重要指标之一。数据采集系统应符合系统安全等保三级的防护要求。

9 实时计算框架技术架构

实时计算数据中心基于高性能的分布式计算框架提供为业务提供低延迟、高吞吐的实时计算任务支持。

9.1 环境要求

实时计算框架主要部署在在类 UNIX 环境上,例如 Linux、Mac OS X 和 Cygwin(对于 Windows),要求集群由一个 master 节点和一个或多个 worker 节点组成。推荐使用的实时计算框架有 Apache Flink、实时计算框架 Streaming 和 Apache Storm 等。

9.2 数据源

实时计算框架应同时支持直接处理和通过消息中间件处理两种数据源接入方式。

9.2.1 直接处理

直接将数据或文件等发送到实时计算框架处理。

9.2.2 通过消息中间件处理

使用消息中间件将发电运行、调度运行、用户用电、电网设备、运营管理、外部等数据发送到实时计算框架,实现对数据的一致性语义的保证。经过实时计算框架处理后,结果数据可用于监报告警、数据分析、异常检测、数据 ETL 等方面的应用。

9.3 存储引擎

本部分可以直接使用机器本机的文件系统。

如果机器本地存储容量存在限制或者系统需要更强的能力,可以使用具有容错的分布式文件系统。

9.4 计算引擎

在多台环境下分布式部署实时计算框架,实时计算框架对输入的数据执行计算任务。

9.4.1 运行进程

实时计算框架运行时应分为作业管理和任务管理两种进程。

9.4.1.1 作业管理器

作业管理器应负责协调应用程序的分布式执行:它决定何时调度下一个 task(或一组 task)、对完成的 task 或执行失败做出反应、协调 checkpoint、并且协调从失败中恢复等等。作业管理器应至少包括以下不同的组件:

1) 资源管理器

资源管理器负责集群中的资源提供、回收、分配。计算引擎应为不同的环境和资源提供者实现了对应的资源管理器,以满足将应用部署中在混合环境的需要。

2) 作业控制器

作业控制器负责管理单个作业图的执行。一个集群中可以同时运行多个作业,每个作业都有自己的作业控制器。

9.4.1.2 任务管理器

任务管理器(也称为 worker)执行作业流的任务(task),并且缓存和交换数据流。一个集群中必须始终至少有一个任务管理器。任务管理器中可以通过资源隔离等方式实现并发处理 task,提高整个集群运行的并行度。

9.4.2 执行流程

实时计算框架执行流程应包括以下环节:

a) 用户提交程序到作业客户端。

- b)作业客户端解析、优化作业，并且将生成的作业提交到作业管理器。
- c)任务管理器运行 task，并上报信息给作业管理器。

9.4.3 容错

作为分布式实时计算引擎，计算引擎应具有强大的容错机制，以确保在出现机器故障或网络分区等不可预知的问题时可以快速自动恢复并依旧能产生准确的计算结果。

9.4.3.1 快照

快照应包括指向每个数据源的指针（例如到文件或消息中间件分区的偏移量）以及每个作业的有状态运算符的状态副本等，该状态副本是处理了数据源偏移位置之前所有的事件后而形成的状态。当集群故障时，可通过快照恢复至错误前的状态的。

9.4.3.2 Checkpoint

为了满足实时计算低延迟的需求，出现故障进行恢复时，计算引擎需要具有快速的错误恢复能力，同时容错副本应该是自动、定期地创建和删除的，不需与用户进行交互，以确保在作业意外失败的情况下可以恢复。

在机器故障时，实时计算框架应可通过 Checkpoint 回退以及重新发送数据源偏移量后数据流从故障中恢复。

9.5 动态扩缩容

实时计算框架应可通过一些监控服务产生的指标，例如：可用 worker、消费延迟、CPU 利用率汇总、吞吐量、延迟等，实现自动扩缩容机制。当上述的这些指标超出或者低于一定的阈值时，增加或者减少 worker 的数量。

9.6 集群架构

实时计算框架应支持集群架构部署。作业管理器是集群中的协调者，负责监控、管理任务管理器的任务执行情况。任务管理器是任务的执行者，通过运行在多个 JVM 进程中的多个线程中来并发执行任务。

10 系统保障

10.1 系统性能提升保障

为实现电力管理系统实时数据采集可用、高效的目标，实时计算框架应实现维表关联，支持消息处理，并满足一致性要求。

10.1.1 维表关联

支持根据实时消息中的某一字段内容，加载外部数据源中的数据，为实时数据进行字段补齐，进而做统计分析工作。

10.1.2 消息系统

电力场景下，接入设备丰富，信息多样，计算复杂，实时计算系统推荐结合消息队列中间件部署，请求数据发送给消息队列之后立即返回，再由实时计算系统中的任务进程作为消息队列的消费者进程从消息队列中获取数据，异步处理接入设备的请求数据，从而能够在数据产生和消费速度不一致的情况下，有效保证整个集群的稳定性。

10.1.3 一致性要求

在实时计算系统，为了保证数据计算的正确，在计算引擎从消息队列中间件种消费数据时，需要保证数据的消费顺序。考虑到处理过程中出错情况，不同的分布式系统有不同一致性状态：

- 1) 最多一次：如果算子处理事件失败，算子将不再尝试该事件。

2) 至少一次：如果算子处理事件失败，算子会再次尝试该处理事件，直到有一次成功。

3) 精确一次：严格地，有且仅处理一次。

本文件的标准场景下，如果需要保证计算的精确一次性，计算引擎的容错需要保证精确一次性，对于计算数据不丢失计算结果也不重复计算。

10.1.4 监报告警框架

在实际生成环境中，本文件推荐通过监控系统来监控服务的健康状况，业务运行情况，关键性能指标等状态。当服务出现问题时可以尽快联系运维人员，保留必要的错误快照信息方便开发人员排障和复盘。

10.2 系统高可用性保障

系统设计时，应通过主从方式（非对称方式）、双机双工方式（互备互援）、集群工作方式（多服务器互备方式）保证电力管理信息系统实时数据采集的高可用性。

通过上述方式，电力管理信息系统实时数据采集的可用性等级应达到5个9，即99.999%。
