



团 体 标 准

T/CES XXX-XXXX

电网实时智能态势评估大数据平台数据 接入规范

Data access specification of big data platform for real-time intelligent situation
assessment of power grid

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国电工技术学会 发布

目 次

前 言	1
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 符号、代号和缩略语	1
5 数据接入方式	1
5.1 实时数据流推送方式	2
5.2 API 获取方式	2
5.3 文件数据传输方式	2
5.4 功能性要求	3
5.5 性能要求	3
5.6 安全性要求	3
6 数据存储方式	3
6.1 结构化数据存储	3
6.2 非结构化数据存储	5
7 数据访问方式	5
7.1 结构化数据访问方式	5
7.2 非结构化数据访问方式	5
附 录 A	7
(规范性)	7
实时数据流格式	7
附 录 B	8
(资料性)	8
数据访问接口说明	8

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》给出的规则起草。

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这些专利的责任。

本文件由中国电工技术学会提出。

本文件由中国电工技术学会标准工作委员会能源智慧化工作组归口。

本文件起草单位：中国电力科学研究院有限公司、中国科学院软件研究所。

本文件主要起草人：刘道伟、乔颖、杨红英、赵高尚、吕先进、蓝海波、宋磊、王凯、卢毅、马鑫晟、张丽、陈勇、杨学涛、刘开欣、杨世通、王世茹。

本文件为首次发布。

电网实时智能态势评估大数据平台数据接入规范

1 范围

本文件规定了电网调度信息接入大数据平台的总体架构、功能要求和技术要求。

本文件适用于电网大数据平台基础量测数据的接入标准,用于支撑电网实时智能态势评估系统的研制与应用。

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 35295-2017	信息技术大数据术语
GB/T 36345-2018	信息技术通用数据导入接口

3 术语和定义

下列术语和定义适用于本文件。

3.1 消息堆积 Message heap

消息发送的速率远远大于消息消费的速率,从而导致大量的消息未被处理存储在缓冲区中,可能会导致数据失去时效性。当大量消息堆积在缓冲区时,可能会导致内存溢出。

3.2 结构化数据 Structured data

能够用数据或统一的结构加以表示的信息,具有高度组织和整齐格式化的数据,包括 PMU、SCADA、GIS、气象数据。

3.3 半结构化数据 Semi-structured data

具有一定的数据格式,数据的结构与内容混合在一起,没有明显的区分,包括电网运行方式、电网计划数据。

3.4 电网实时智能态势评估 Intelligent situation assessment of power grid

基于海量的电网量测、仿真数据,采用先进的大数据、人工智能技术实现的电网在线态势评估。

4 符号、代号和缩略语

下列符号、代号和缩略语适用于本文件。

PMU: 同步相量测量单元 (Phasor Measurement Unit)
RTU: 远程终端单元 (Remote Terminal Units)
SCADA: 数据采集与监视控制系统 (Supervisory Control And Data Acquisition)
GIS: 地理信息系统 (Geographic Information System)
API: 应用程序编程接口 (Application Programming Interface)
SFTP: 安全文件传输协议 (Secret File Transfer Protocol)
ID: 标识 (Identifier)
JSON: JavaScript 对象表示格式 (JavaScript Object Notation)
SSD: 固态硬盘 (Solid State Drives)
HDD: 硬盘驱动器 (Hard Disk Drive)
HTTPS: 超文本传输安全协议 (Hypertext Transfer Protocol Secure)

5 数据接入方式

5.1 实时数据流推送方式

实时数据流推送方式用于接入实时量测数据，包括 PMU、SCADA 数据。实时数据流推送方式可保证此类量测数据能够实时的接入到实时智能态势评估大数据平台（简称大数据平台，后续如无特殊说明，大数据平台均指实时智能态势评估大数据平台），如图 2 所示。



图 2 实时数据流推送方式

具体步骤如下：

- 1) 定义数据流格式，将各类量测数据按照所定义的格式转换成数据流，所定义的数据流格式见附录 A；
- 2) 将数据流推送到分布式消息队列中；
- 3) 大数据平台实时从分布式消息队列中获取数据；
- 4) 将所获取的数据流解析为相应的大数据平台存储格式，并存储在大数据平台，其相应存储格式与存储方式见 6.1。

5.2 API 获取方式

API 获取方式用于接入更新频率低的结构化数据，包括 GIS、气象数据。GIS 与气象数据存储在关系型数据库中。可通过周期性调用数据库提供的数据库接口，获取相应的 GIS、气象数据，如图 3 所示。

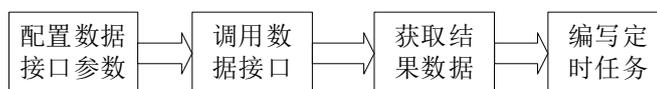


图 3 API 获取数据方式

具体步骤如下：

- 1) 配置数据接口所需参数，例如添加动态链接库路径；
- 2) 通过调用数据库提供的数据库接口获取 GIS 数据与气象数据；
- 3) 将 2) 中的返回结果格式转换成相应的大数据平台存储格式，并存储在大数据平台，其相应存储格式与存储方式见 6.1；
- 4) 编写定时任务脚本，周期性执行上述 2-3 的步骤。

5.3 文件数据传输方式

文件数据传输方式用于接入非结构化文本数据，包括电网运行方式数据、生产计划数据。电网运行方式数据与生产计划数据均以文本文件的形式存储在调度自动化系统中。可采用 SFTP 协议，获取这些文本文件，其传输方式如图 4 所示。

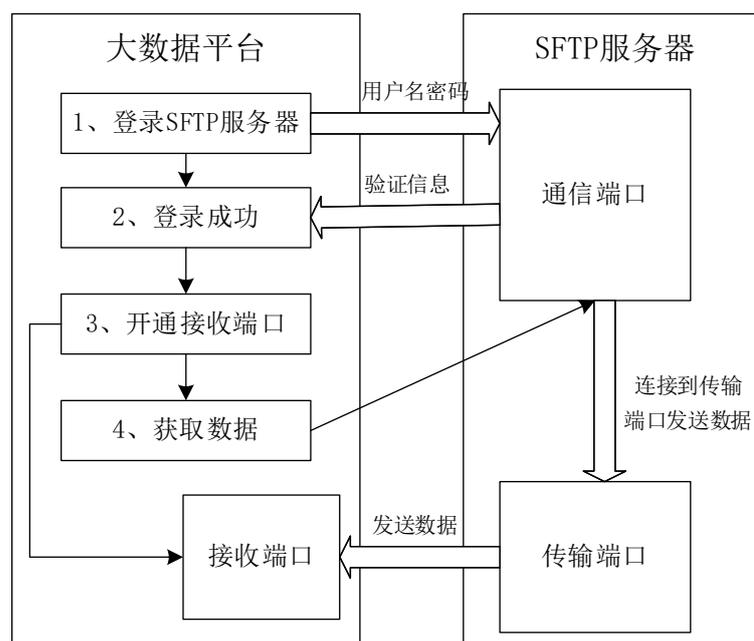


图 4 文件数据传输方式

具体步骤如下：

- 1) 驻留在大数据平台的接入程序通过用户名密码登录 SFTP 服务器；
- 2) 获取文件数据存储位置；
- 3) 定时从 SFTP 服务器下载文件数据；
- 4) 退出 SFTP 服务器。

5.4 功能性要求

电网调度数据接入大数据平台的过程中，需要保证数据接入的实时性、完整性与接入程序的稳定性，具体功能性要求如下：

- 1) 从分布式消息队列获取数据流时需要保证消费数据的速度，保证消息队列不会出现消息堆积，实现数据实时接入；
- 2) 制定数据序列化方案，利用 Protocol Buffer 数据协议保证接入数据的正确性与完整性；
- 3) 对数据接入程序的读取数据、数据格式转换过程进行监控，保证数据接入程序稳定运行。

5.5 性能要求

针对省级以上电网，大数据平台接入电网量测、仿真数据性能需满足以下要求：

- 1) PMU 数据接入速度需达到每秒接入五万以上测点数据；
- 2) SCADA 数据接入速度需达到每秒接入一个时间断面下所有遥信、遥测数据；
- 3) 文件型数据（电网运行方式数据、生产计划数据）接入速度需达到每秒接入一个时间断面数据。

上述功能性要求与性能要求针对大数据平台与数据源系统均在生产管理区（安全区 III）。

5.6 安全性要求

针对省级以上电网，大数据平台接入电网量测、仿真数据安全性需满足以下要求：

- 1) 当数据接入程序出现异常时，应及时排查原因，超过一定时长未解决，则需终止接入程序，以免影响现有系统稳定运行；
- 2) 通过 API 方式获取数据时，应禁止对数据库进行添加、修改、删除操作。

6 数据存储方式

6.1 结构化数据存储

6.1.1 PMU 数据存储

将 PMU 数据分为在线数据与历史数据，将当前时刻前五分钟内的 PMU 数据作为在线数据，存储在分布式内存数据库中；其余时刻数据作为历史数据，存储在分布式列式数据库中。

在线 PMU 数据采用键值对的形式进行存储。如表 1 所示，键由 PMU 时标、PMU 测点地址以及每秒所含数据帧数组成，值由 N 个数据帧对应的数据组成。

表 1 在线 PMU 数据存储格式

Key (键)	Value (值)
时标_PMU 测点地址_数据帧数	第一帧数据, 第二帧数据, ..., 第 N 帧数据

历史 PMU 数据以行列式进行存储。如表 2 所示，行键由 PMU 测点地址与 PMU 时标组成，列簇表示所有数据列的集合，数据列 frame 表示每秒所含数据帧数，value1, value2, ...value N 分别为每一帧数据。

表 2 历史 PMU 数据存储格式

RowKey (行键)	Column Family (列簇)				
PMU 测点地址_时标	frame	value1	value2	value3	valueN
	每秒数据帧数	第一帧数据	第二帧数据	第三帧数据	第 N 帧数据

6.1.2 SCADA 数据存储

将 SCADA 数据分为电网模型数据与电网运行数据，电网模型数据主要包括设备拓扑、设备静态参数信息，存储在关系型数据库中；电网运行数据包括设备遥信信息、遥测信息，将当前时刻前一小时内的电网运行数据作为实时运行数据，存储在分布式内存数据库中，将其它时刻电网运行数据作为历史运行数据存储在分布式列式数据库中。

电网模型数据以设备类型进行划分，采用数据表的形式存储在关系型数据库中，同一类型设备的模型数据存储在一张数据表中。以母线型设备为例，其模型数据的数据表共包含九个字段，各字段含义如表 3 所示。

表 3 母线型设备的模型数据存储格式

id	Name	st id	vol type	run state	node id	bs type	vmax	vmin
设备 ID	设备名称	厂站 ID	电压等级	运行状态	物理连接节点	母线类型	电压上限	电压下限

电网实时运行数据以键值对的形式存储在分布式内存数据库中。如表 4 所示，键由 SCADAId (SCADA 量测点唯一标识) 与时标组成，值由对应的量测值组成。

表 4 实时运行数据存储格式

Key (键)	Value (值)
SCADAId_时标	量测值

电网历史运行数据以行列式进行存储。如表 5 所示，行键由 SCADAId 与时标组成，列簇表示所有数据列的集合，包括四个数据列，分别为有功功率、无功功率、电压、电流的具体量测值。

表 5 历史运行数据存储格式

RowKey (行键)	Column Family (列簇)			
SCADAId_时标	p	q	v	i
	有功功率	无功功率	电压	电流

6.1.3 GIS 数据存储

采用关系型数据库存储 GIS 数据，数据表中包含五个字段，其字段含义如表 6 所示。

表 6 GIS 数据存储格式

id	stationName	type	lng	lat
序号	厂站名称	厂站类型	经度	纬度

其中厂站类型约定为电厂、变电站、开关站以及换流站四类。

6.1.4 气象数据存储

采用关系型数据库存储气象数据，数据表中包含七个字段，其字段含义如表 7 所示。

表 7 气象数据存储格式

id	stationId	stationName	weatherDate	weatherSpeed	ltemp	htemp
序号	站号	站名	天气时间	风速	最低温度	最高温度

6.2 非结构化数据存储

采用分布式文件系统存储非结构化数据，包括电网运行方式、生产计划数据，具体功能如下：

- 1) 采用分布式文件存储，支持文件多副本，分块存储；
- 2) 根据数据时标，按照“年/月/日”的层次结构设计文件目录，采用数据时标作为文件名称；
- 3) 根据数据使用频率进行冷热数据分区，将热数据存储于 SSD 中，冷数据存储于 HDD 中，提高数据访问速度。

7 数据访问方式

7.1 结构化数据访问方式

结构化数据包括 PMU 数据、SCADA 数据、气象数据以及 GIS 数据。根据上层业务需求，提供结构化数据访问接口，支持大数据平台结构化数据存取操作，结构化数据访问接口说明见附录 B。如图 5 所示，结构化数据访问接口实现方式如下：

- 1) 大数据平台根据查询需求，创建数据索引，形成分布式检索引擎；
- 2) 大数据平台获取查询条件，从分布式检索引擎中获取满足条件的数据 id；
- 3) 大数据平台利用数据 id 从底层数据库中获取满足条件的数据；
- 4) 大数据平台将满足条件的数据转化成 JSON 格式返回给上层业务应用。

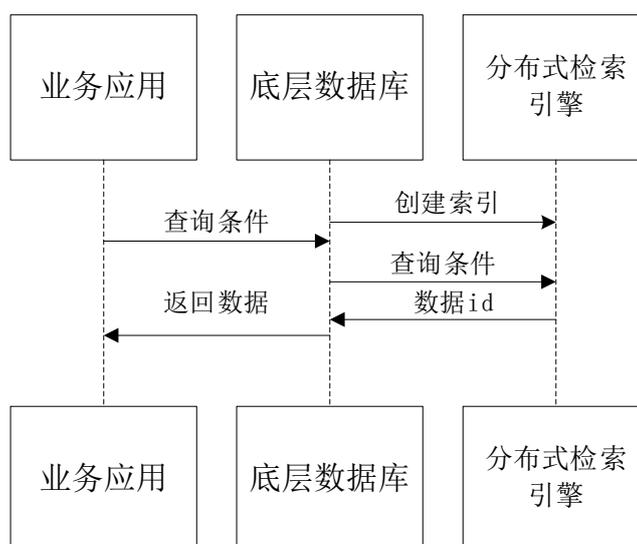


图 5 结构化数据访问接口实现方式

7.2 非结构化数据访问方式

非结构化数据主要包括电网运行方式、生产计划等文本型数据。根据上层业务需求，提供非结构化数据访问接口，支持上层业务应用直接获取文本型数据。非结构化数据访问接口说明见附录 B。如图 6 所示，非结构化数据访问接口的实现方式如下：

- 1) 根据查询条件，从中筛选出待查询数据的时间范围；
- 2) 根据时间范围定位到满足条件的数据文件所在目录与文件名称；
- 3) 将满足条件的数据文件以字符串的形式返回给上层业务应用。

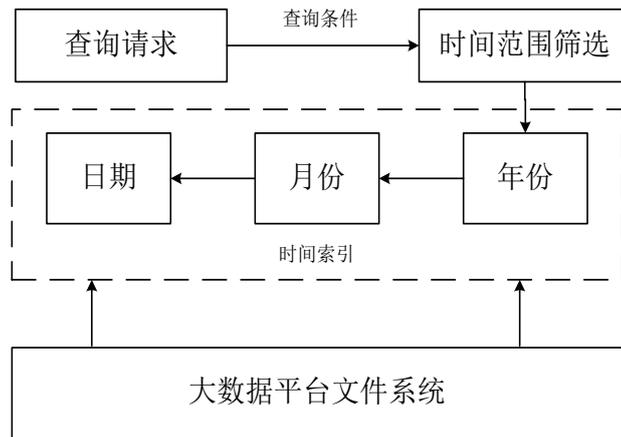


图 6 非结构化数据访问接口实现方式

附录 A

(规范性)

实时数据流格式

A.1 PMU 数据流结构定义应符合表 A.1。

表 A.1 PMU 数据流结构

结构名称	字段名称	数据类型	描述
消息头	package_type	int (整数类型)	存储数据的类型 (遥测值、遥信值)
	data_num	int (整数类型)	传输点的个数
	msg_per_sec	int (整数类型)	每秒设定的子站上送报文的总帧数
	second	time t (时间类型)	当前报文的子站上送时间
消息体	mes	char[] (字符数组)	遥测/遥信值, 前 4 字节为 PMU 测点地址

A.2 SCADA 数据流结构定义应符合表 A.2。

表 A.2 SCADA 数据流结构

数据类型	字段名称	数据类型	描述
遥信	keyid	long (长整型)	关键字, SCADA 量测点唯一标识
	value	char (字符类型)	遥信值
	status	int (整数类型)	遥信状态
遥测	keyid	long (长整型)	关键字, SCADA 量测点唯一标识
	value	float (浮点类型)	遥测值
	status	int (整数类型)	遥测状态

附录 B

(资料性)

数据访问接口说明

B.1 结构化数据访问接口说明见表 B.1.

表 B.1 结构化数据访问接口

数据接口名称	请求方式	请求参数	返回格式 (JSON 格式)
PMU 数据访问接口	POST	startTime: 开始时间	status: 请求状态 data: PMU 数据
		endTime: 结束时间	
		devname: 设备名称	
		paraName: 属性名称	
SCADA 数据访问接口	POST	startTime: 开始时间	status: 请求状态 data: SCADA 数据
		endTime: 结束时间	
		devName: 设备名称	
		paraName: 属性名称	
GIS 数据访问接口	POST	station: 厂站名称	status: 请求状态 data: GIS 数据
气接口象数据访问	POST	time: 时间	status: 请求状态 data: 气象数据
		station: 站点名称	

B.2 非结构化数据访问接口说明见表 B.2

表 B.2 非结构化数据访问接口

数据接口名称	请求方式	请求参数	返回格式 (JSON 格式)
电网运行方式数据访问接口	POST	startTime: 开始时间	status: 请求状态 data: 电网运行方式数据字符串
		endTime: 结束时间	
生产计划数据访问接口	POST	startTime: 开始时间	status: 请求状态 data: 生产计划数据字符串
		endTime: 结束时间	

注: POST 是数据查询的一种专用方式