

# 基于大数据技术构建输变电设备状态监测数据模型研究

郭宝<sup>1</sup> 邵进<sup>2</sup>

<sup>1</sup>(南瑞集团有限公司国网电力科学研究院 江苏 南京 210000)

<sup>2</sup>(国家电网公司运维检修部 北京 100000)

**摘要** 在输变电设备运行数据的监测和采集过程中,由于系统建设的阶段性、技术性、资金等因素影响,导致电力企业在系统建设过程中积累了大量采用不同方式存储的电网设备运行实时数据,包括设备运行数据、环境数据、气象数据等。从简单的文件数据库到复杂的网络数据库,构成电力企业的异构数据源。这些异构数据由于数据模型、存储方式和传输协议存在差异,导致信息共享困难,严重影响输变电设备运行监测数据的分析和应用。在此背景下,利用大数据存储技术,攻克异构数据源之间数据共享和业务协同的技术难点,为输变电设备评价、状态评价和运维决策等应用提供数据支撑,提升设备状态管理信息化水平,强化电网运行的可靠性和安全性。

**关键词** 设备模型 数据共享 状态监测 大数据

中图分类号 TP31 文献标识码 A DOI:10.3969/j.issn.1000-386x.2019.02.024

## CONSTRUCTION OF MONITORING DATA MODEL FOR TRANSMISSION AND TRANSFORMATION EQUIPMENT BASED ON BIG DATA TECHNOLOGY

Guo Bao<sup>1</sup> Shao Jin<sup>2</sup>

<sup>1</sup>(NARI Group Corporation/State Grid Electric Power Research Institute, Nanjing 210000, Jiangsu, China)

<sup>2</sup>(Department of Maintenance and Repair, State Grid Corporation, Beijing 100000, China)

**Abstract** In the process of monitoring and collecting the operation data of power transmission and transformation equipment, due to the influence of phases, technologies and funds of system construction, a large number of real-time data of power grid equipment operation stored in different ways have been accumulated in the process of system construction, including equipment operation data, environmental data and meteorological data. From a simple file database to a complex network database, it constitutes the heterogeneous data source of an enterprise. Due to differences in data models, storage methods and transmission protocols, these heterogeneous data make it difficult to share information, which seriously influences the analysis and application of monitoring data of power transmission and transformation equipment. Under this background, the data storage technology of big data was used to overcome the technical difficulties of data sharing and business collaboration among heterogeneous data sources. It provides data support for the application of power transmission and transformation equipment status evaluation and operation and maintenance decision-making. It improves the information level of equipment state management, and strengthens the reliability and safety of power grid operation.

**Keywords** Equipment model Data sharing Status monitoring Big data

## 0 引言

电网设备安全可靠运行是整个电力系统安全的基础,某一局部的故障常常会影响电网安全运行,给用

户、企业和社会造成不可预估的损失。在当前电力需求每年高速增长和电力企业商业化运营环境下,输变电设备的健康状态和维护成本直接影响设备运行可靠性、企业效益和市场竞争力。因此,以可靠性为中心的维修策略(RCM)是全面掌控输变电设备运行状态,提

升运维精益化管理水平的重要手段。目前,大多数网省公司成立了设备状态评价中心,提升了输变电设备状态评估,改进了设备故障分析诊断,加强对设备全生命周期成本分析水平。

在电力企业信息化建设过程中,各业务系统建设和实施过程中,由于系统建设阶段和技术应用因素等影响,导致电力企业积累了大量业务数据,从简单的文件数据库到复杂的网络数据库,它们构成了电力企业的异构数据源。虽然这些数据能够满足各专业应用,但是在许多情况下常常为了一个决策分析,必须访问多个业务应用系统。这些错综复杂的应用,涉及到不同系统、不同平台和不同应用,甚至涉及到不同企业间的异构分布式数据交互。因此,为了在这些异构数据环境中有效访问这些数据,并从中获取所需的数据信息,提出数据整合概念,而数据整合源于数据集成技术。

20 世纪 70 年代中期开始,数据集成技术在数据集成深度和广度上,都在不断地扩大。异构数据库的集成技术通常有以下几种:数据迁移和转换技术、中间件技术、联邦数据库技术。随着新技术的迅猛发展,利用“大数据”技术整合数据,形成有效的数据架构,实现跨业务平台整合不同来源、结构化和非结构化的数据,从而帮助识别数据、消除冗余、优化文件系统,为数据分析提供基础。同时通过引进“大数据”技术,对数据进行有效、充分地整理和分析,减少或压缩无价值的数 据,提高有效数据的利用价值,提升精益化管理水平,促进新业务形态发展。

## 1 数据源分析

### 1.1 异构多源数据分析

输变电设备的运行状态评估流程如图 1 所示。主要通过建立输变电设备健康档案,并在健康档案的数据基础上开展输变电设备状态评估、运维策略优化和监测预警等应用。

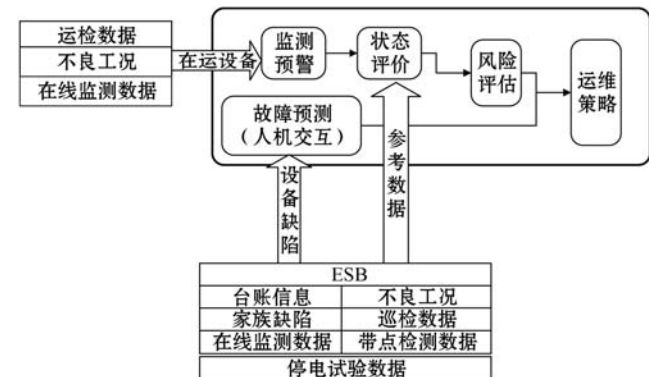


图 1 输变电设备状态评估

由于输变电设备的健康档案数据来源多样,有的来源于电网设备运维精益管理信息系统;有的来源于输变电状态监测系统;有的来源于电力自动化系统(SCADA);有的记录在设备检测报告中;有的记录在例行试验表格中,这些数据分散在多个源头,数据存储形式、类别、格式均不同。因此,为了实现输变电设备监测数据的集成共享应用,必须解决多源异构数据统一建模。

目前,解决多源异构系统数据集成的方法很多,如:数据库系统、数据库迁移和转换、建立数据仓库和中间件集成。然而,从当前系统集成应用效果分析比较,这些方案未能有效地解决业务协调和数据集成共享。在此背景下,我们提出一种基于大数据平台的异构多源数据集成和存储,通过构建输变电设备状态监测数据交互模型,存储设备档案信息,用于设备状态评估和模型诊断,进一步深度挖掘输变电设备监测数据价值,分析电网设备运行状态与环境、时间、检修、实验等因素的关联性,为电网设备可靠运行提供技术支撑手段。

### 1.2 异构多源数据接入

输变电状态监测接入数据类型主要分为实时和非实时两种。接入方式如图 2 所示。

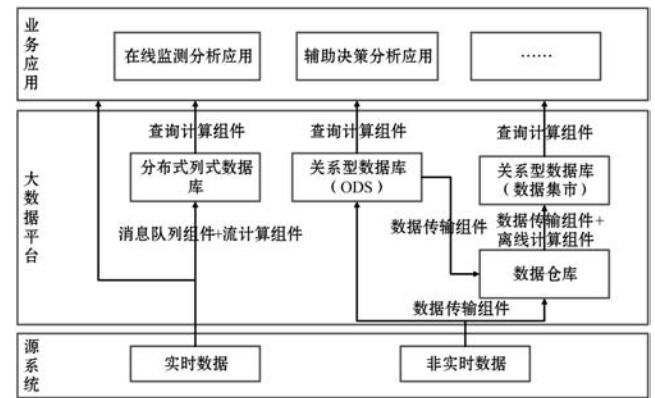


图 2 异构多源数据接入

非实时性数据主要包括设备信息、检修信息、故障信息、技术设备缺陷信息等,这类数据主要通过大数据平台数据传输组件,将源系统静态数据抽取至数据仓库,事务型数据抽取至关系型数据库中,按需组装数据,用于前端应用展示,或共享至其他业务系统,并定期进行归仓。数据仓库中的数据按需进行批量计算,将结果数据存放至关系型数据库(数据集市),用于前端应用展示。

输变电设备的实时数据包括有功、无功、电流、电压及开关状态等,这类数据主要通过大数据平台消息队列+流计算组件,接收源系统实时数据并进行计算,用于前端应用实时展示。源明细数据及计算结果数据

按需存储至非关系型数据库,用于前端应用展示或数据分析挖掘。

## 2 数据模型设计

### 2.1 资源区域划分

利用大数据平台数据资源存储的物理模型管理功能,定义源数据、共享数据、分析数据存储的结构,并划分不同的数据存储区域,按照数据应用需求进行设计。设计内容包括数据缓冲区、数据视图区和数据仓库的存储模型。数据存储区域划分主要有以下几种:

(1) 数据缓冲区 数据缓冲区是输变电状态监测数据进入大数据平台的统一数据区域,源业务系统数据首先存储于数据缓冲区,缓冲区内的数据保持与源系统一致。数据缓冲区的数据库采用 MYSQL,使用 SQL 标准语言访问数据表,大数据平台存储区管理为每个业务系统分别创建了一个数据缓冲区。该区的模型采用帖近源的方式设计,模型与源业务系统数据模型保持一致。

(2) 数据视图区 数据视图区主要存放大数据平台资源整合和规范后的数据,对于共享或快速应用的输变电状态监测实时数据,统一进入数据视图区。

该区模型采用扁平化方法设计,遵循国网 SG-CIM 模型标准,结合输变电设备监测采集数据的需求,按照大数据平台 MYSQL 存储架构特性,设计形成统一数据视图存储模型。数据视图存储模型统一遵循 SG-CIM 模型的表、字段命名规范和主题域划分等要求。

(3) 数据仓库区 数据仓库存放输变电状态监测的历史数据,对所有接入大数据平台的数据进行归仓管理,对于需要挖掘分析的数据统一基于数据仓库对外提供,数据仓库采用 HIVE,数据库用户只有一个。

输变电设备监测的数据仓库存储模型遵循 SG-CIM 标准规范,按照大数据平台的 HVIE 数据特性存储,并结合数据应用需求,采用扁平化设计。

### 2.2 存储模型设计

当前各专业应用系统通过消息队列+流计算方式,实时接入大数据平台的实时采集数据,按 SG-CIM 规范存储,输变电设备状态监测系统、用电信息采集系统、配电自动化系统、电能质量监测系统等。实现数据区域直接访问,获取电力测点数据,进行数据分析和应用。

实时采集数据的存储主要基于非关系型数据库(HBase)之上构建,数据模型按照<测点名、量测值、质量码、时间戳>四元组方式组织,这四元组是输变电

设备监测数据采集量测数据的基本特征。

非关系型数据库的四维列式存储遵循 SG-CIM 规范,数据存储模型示例如表 1 所示:① row-key 为行键;② 列族(column-family);③ 列限定符(column-qualifier);④ 版本(time-version)。

表 1 列式存储模型

行建 ①	② 列族-info		
	NAME ③	EMAIL	PASSWORD
GGrandpaID	张大力	zhangdali@126.com	123
HMS_Surprise	王宝强	wangbaoqiang@126.com	123
StrDoyle	李石敏	lishimin@126.com	456
TheRealMT	张丽茹	zhangliru@126.com	④ 111

每个数据单元保留多个时间版本,由被插入表中的时间戳来记录。非关系型数据库模型设计关键步骤如下:

(1) 列族采用单字母命名,同时列和列族的所属按数据使用频度和使用方式归类,将实时和非实时查询的列,分别归类于不同的列族,存储多个属性和字段,并用分隔符隔开。

(2) 行键长度一般以 3 至 10 个字节为宜,不宜超过 16 个字节。散列字段作为行键的高位,低位放置时间字段,这样将提高数据均衡分布在每个分区实现负载均衡的概率。

(3) 行键必须保证唯一,否则会导致数据被覆盖。

(4) 表将根据数据场景需求进行设计,对具有关联关系的实体按需组合设计形成一张表。

(5) 表按需分为 5 至 20 个区,根据二进制逐位比较,决定采集数据具体的存储方式。

### 2.3 数据访问设计

通过对实时监测数据、视频信息数据和多元离线数据的集成整合存储,由大数据平台提供统一数据访问服务,实现设备集中监控、设备状态评价和故障诊断等应用。输变电设备运行监测数据访问服务分为两类:实时数据访问接口和非实时数据访问接口。实时监测数据的访问设计,遵照 kafka 消息队列开发规范设计。非实时监测数据访问接口,通过 JDBC 连接数据库访问监测数据,无需访问数据仓库,而是直接通过数据缓冲区或统一数据视图区获取:

(1) 业务应用通过大数据平台提供的访问用户名(该用户只有读取权限),直接访问数据仓储读取数据,主要适用于通过数据视图区读取数据的业务应用场景。

(2) 大数据平台将为业务系统访问监测数据提供若干服务接口,包括地址访问服务、数据读取服务、数

据操作服务等。如表 2 所示。

表 2 数据访问服务

名称	分布式列式数据库接口		
包名	com. sgcc. bigdata. api		
类名	GdbaseClient		
方法名	中文名称	英文名称	描述
	模糊查询	rowFilterScan	提供分布式列式数据库更近所列值进行模糊查询的功能
	表存在核查	checkTable	提供分布式列式数据库根据所给表名查询是否已存在的功能
	删除单条记录	deleteRecord	提供分布式列表数据库根据所给行健值删除单条记录功能
	批量删除	deleteRecords	提供分布式列式数据库批量删除的功能
	删除表	deleteTable	提供分布式列式数据库删除表的功能
	批量插入	insertRecords	提供分布式列式数据库批量数据写入的功能
	插入单条记录	insertRecord	提供分布式列式数据库单条记录写入的功能
	行健查询	getOneRecord	提供分布式列式数据库根据所给行健值查询结果集的功能
	全表扫描	getAllRecord	提供分布式列式数据库全表扫描结果集功能
	创建表	createTable	提供分布式列式数据库创建表的功能

通过大数据平台提供的数据库接口,实现数据库初始化和数据查询、更新和删除等操作。实例如下:

```
public class JdbcConnDemo {
    public static String url = "jdbc: hive2://10. 225. 10. 19:
9001/default";
    public static String name = "SCYX";
    public static String Password = " ";
    public static void main(String[] args) throws Exception {
        //调用 Jdbc 获得数据库连接
        newconn = GddwUtil. getConn(url, name, Password);
        if( newconn != null) {
            System. out. println("Connection is successful!");
        } else {
            System. out. println("Connection failed!");
        }
    }
}

public class CreateTableDemo {
    public static void main(String[] args) throws Exception {
        GdbaseClient gdbaseClient = new GdbaseClient();
        //传入参数表名
        String tableName = "gdbase_tableA";
```

```
//列族
String[] columnFamily = {"bguo"};
//创建表
gdbaseClient. createTable(tableName, columnFamily);
}

public static void main(String[] args) throws Exception {
    GdbaseClient gdbaseClient = new GdbaseClient();
    //传入数据表参数
    String tableName = "gdbase_tableA";
    //全表扫描
    List < GCell > results = gdbaseClient. getAllRecord( ta-
bleName);
}
}
```

### 3 结 语

基于大数据平台构建输变电设备状态监测运行数据交互模型,为各专业管理提供数据共享和业务协同服务,以规范专业数据的有序存储、流转和处理计算分析等,确保数据有效可用。在此基础上构建输变电设备健康档案信息,依据输变电设备运行工况的当前和历史数据,利用“浴盆曲线”、“指纹分析法”和“趋势预测法”等,开展输变电设备资产数据、环境数据和运行工况数据之间关联因素分析,建立输变电设备状态分析模型,为设备状态评价和运维决策提供数据支撑。

### 参 考 文 献

- [1] 严英杰, 盛戈皞, 陈玉峰, 等. 基于大数据分析的输变电设备状态数据异常检测方法[J]. 中国电机工程学报, 2015, 35(1):52-59.
- [2] 张博文, 阎春雨, 毕建刚, 等. 基于大数据的输变电设备状态预警系统架构研究[J]. 电力信息与通信技术, 2016, 14(12): 26-32.
- [3] 郭创新, 熊世旺, 张行, 等. 输变电设备全景信息的多源异构数据聚合方法[J]. 高电压技术, 2015, 41(12): 3888-3894.
- [4] 李为, 周琰, 成永强, 等. 基于 CIM 扩展的输变电在线监测系统模型[J]. 电气技术, 2017, 18(1):77-82.
- [5] 罗怿, 尹立群, 胡军, 等. 输变电设备故障大数据可视化分析方法研究[J]. 电力信息与通信技术, 2016, 14(7): 12-16.
- [6] 王德文, 李静芳. 变电设备状态监测大数据的查询优化方法[J]. 电力系统自动化, 2017, 41(2):165-172.
- [7] 周广, 闫丹凤, 许光可, 等. 大数据在输变电设备状态评估中的研究[J]. 软件, 2016, 37(1):9-13.