

DOI:10.13296/j.1001-1609.hva.2019.04.023

## 电网雷电大数据采集系统研发

朱 晔<sup>1,2,3</sup>, 姜志博<sup>1,2,3</sup>, 田 浩<sup>1,2,3</sup>, 陶汉涛<sup>1,2,3</sup>, 吴大伟<sup>1,2,3</sup>,  
张 磊<sup>1,2,3</sup>, 何 君<sup>1,2,3</sup>

(1. 南瑞集团有限公司, 南京 211106; 2. 国网电力科学研究院武汉南瑞有限责任公司, 武汉 430074;  
3. 电网雷击风险预防湖北省重点实验室, 武汉 430074)

**摘要:** 随着中国电网雷电监测预警网中雷电相关探测设备规模不断扩大,为提升雷电数据采集的实时性、完整性和准确性,研发了电网雷电大数据采集系统。阐述了雷电大数据的产生,分析了现有雷电数据采集方案存在的问题,说明了基于分布式集群的系统关键技术方案,介绍了雷电大数据采集系统具体实现的功能,并选择某省电网进行系统的部署及应用。应用效果表明,提出的电网雷电大数据采集系统能显著提升数据采集性能,同时具备全方位监控、智能化运维等特点,为雷电数据采集设备提供了统一的、稳定的、可扩展的数据接入平台,对电网雷电监测预警网的安全稳定运行具有重要意义。

**关键词:** 数据采集; 分布式; 集群; 雷电大数据; 智能运维

## Research and Development of Lightning Big Data Acquisition System in Power Grid

ZHU Ye<sup>1,2,3</sup>, JIANG Zhibo<sup>1,2,3</sup>, TIAN Hao<sup>1,2,3</sup>, TAO Hantao<sup>1,2,3</sup>, WU Dawei<sup>1,2,3</sup>,  
ZHANG Lei<sup>1,2,3</sup>, HE Jun<sup>1,2,3</sup>

(1. NARI Group Corporation Ltd., Nanjing 211106, China; 2. Wuhan NARI Limited Liability Company, State Grid Electric Power Research Institute, Wuhan 430074, China; 3. Hubei Key Laboratory of Power Grid Lightning Risk Prevention, Wuhan 430074, China)

**Abstract:** With the continuous expansion of lightning related detection equipment in the lightning detection and warning network of China's power grid, in order to improve the real-time, integrity and accuracy of lightning data acquisition, a lightning big data acquisition system is proposed and developed. In this paper, it describes the generation of lightning big data, then analyzes the existing problems of the existing lightning data acquisition scheme, then shows the key technical scheme of the system based on distributed cluster technology, then explains the realized function of lightning big data acquisition system, and then choose a province power grid for system deployment and application. The effect of application shows that the lightning big data acquisition system proposed in this paper can significantly improve the data acquisition performance, and it has the characteristics of omnidirectional monitoring and smart operation. It provides a unified, stable and extensible data access platform for lightning detection equipment, and it is of great significance to the safe and stable operation of lightning detection and warning network.

**Key words:** data collection; distribution; cluster; lightning big data; smart operation

## 0 引言

随着中国电网雷电监测预警网的不断完善,雷电探测设备已基本覆盖全国各省市并逐渐增加,

这些雷电探测设备每年获取的雷电数据量达到TB级<sup>[1-4]</sup>。为实现全国海量雷电数据的接入,国网雷电监测与防护技术实验室团队研发了基于mina框架的雷电数据采集系统并应用至今,然而实际运行中

收稿日期:2018-09-10; 修回日期:2018-11-05

基金项目:国家自然科学基金项目(51777037);国家电网公司科技指南项目(524625170011)。

Project Supported by National Natural Science Foundation of China(51777037), Science and Technology Guide Project of SGCC (524625170011).

雷电短时集中爆发时性能有所下降。当中心服务器或数据通道发生问题时,整个雷电监测预警系统将无法正常运行。此外,随着雷电数据采集装置规模扩大以及种类增加,不仅存在由数据短时集中高并发引起的数据处理不及时、动态负载不均衡和数据丢失等问题,也会面临多类型雷电采集装置的数据接入不便于统一管理的情况。现阶段,一般采用人工进行数据补齐、定时维护和分散管理,难以满足雷电监测预警系统日常的运维需求。因此,研究大数据在智能电网中的技术发展<sup>[5-6]</sup>,引进新技术进行高并发数据处理,开发统一的可扩展的雷电大数据采集平台具有十分重要的意义。

分布式集群技术具有优化网络访问资源、高可用性等特点<sup>[7-9]</sup>,结合其在负载均衡、数据导入、数据转发、实时计算、数据存储、数据交换等方面的技术优势,研发电网雷电大数据采集系统,能够较好地提升高并发雷电数据处理的实时性、完整性和准确性。同时,引入针对分布式集群系统的监控技术,在发生问题前及时预警,实现智能化运维,能更好的保证采集系统安全稳定运行。

文中通过阐述雷电大数据的产生,分析了现有雷电数据采集方案存在的问题;在此背景下,说明了基于分布式集群的系统关键技术方案,完成了电网雷电大数据采集系统总体架构的设计,并介绍了采集系统具体实现的功能;最后,选择某省域电网进行系统的部署及应用效果验证,结果表明:研发的电网雷电大数据采集系统能显著提升数据采集性能。

## 1 雷电大数据的产生

雷电的发生,伴随着强烈的电磁场变化。这些电磁场信号为雷电探测提供了依据。雷电探测技术是雷电科学的重要研究领域<sup>[10]</sup>。我国雷电监测预警网采用的雷电探测设备主要有:高精度低频雷电探测仪、多普勒天气雷达、地面电场仪<sup>[11-12]</sup>,新一代探测设备高频雷电探测站及雷声探测站也处于研发并部署应用中<sup>[13]</sup>。这些探测设备采集的雷电数据物理量包含探测站的运行状态及雷电发生时间、角度、电强度、偏移值、陡度、大气电场值、雷电峰值点、雷声方位角等多种类参数,截至2017年底,中国已建各类雷电探测站点超过1 000个,已基本覆盖全国,其分布见图1。

每个雷电探测站点24 h不间断地监控空间雷电波信息,数据量巨大,每秒达到30字节,全部探测站点每年产生的原始数据约1 TB。由于这些原始



图1 雷电监测预警网雷电探测设备分布图  
(地图所示为局部区域)

Fig. 1 Distribution map of lightning detection equipment for lightning detection and warning network (parts of Chinese map)

数据属于不可再生数据,且雷电相关统计分析结果数据均基于此,因而保证雷电数据采集系统的稳定运行至关重要。

## 2 现有雷电数据采集方案

现有雷电数据采集系统基于mina架构<sup>[14-16]</sup>进行数据传输,其结构见图2。

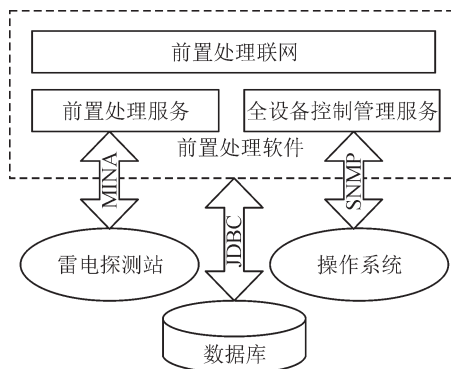


图2 现有雷电数据采集系统结构

Fig. 2 Structure of existing lightning data acquisition system

其中,前置处理服务模块负责与探测站进行通信及数据处理;全设备控制管理服务模块负责与操作系统进行通信及数据处理。接收到的本地探测站主动发送的状态信息及雷电数据通过前置处理服务模块解析处理后存数据库。

前置处理联网模块采用“总中心站—分中心站—子中心站”3层结构进行多层级组网。雷电探测站将数据传至子中心站,子中心站再将数据转发至分中心站,最后由总中心站负责所有分中心站数据的汇总。

然而在实际应用中,常发生数据处理不及时、动态负载不均衡、数据丢失等问题。运行结果显

示,国网已建成投运的雷电探测站在2018年4月的数据接入在线率不到80%,为运维工作带来极大的挑战。同时雷电监测与预警的数据采集系统各自分散,不利于统一管理。

### 3 系统关键技术方案

雷电大数据采集系统采用分布式集群结构,在系统中有多个注册中心及工作节点。当某一服务器宕机时,正常运行节点会接替宕机的工作节点。由于处理能力分布在多个服务器中,可以实现比现有采集方案更快的数据处理速度,实现多物理量雷电大数据同时采集,并保证数据的完整性。以下对融合分布式集群技术的雷电大数据采集系统关键技术方案进行详细介绍。

#### 3.1 负载均衡

负载均衡是一种扩展服务器带宽、增加吞吐量、加强网络数据处理能力以及提高网络传输速率的灵活性技术。LVS(linux virtual server)采用3层结构,由负载调度器、服务器池、共享存储组成,具有良好的可靠性、扩展性和操作性,且相较于高成本的商用F5硬件负载设备,其负载均衡技术廉价、易用,因此被广泛地应用在大数据平台的负载处理环节。

在雷电大数据采集系统数据接入服务中,LVS和网关服务部署在同一个局域网内,二者IP属于相同的网段,且每个网关所在服务器的配置相同,故在负载均衡中主要使用LVS的DR(direct routing)模式和LC(least connections)调度算法<sup>[17-18]</sup>。LVS基于IP和端口实现,对于不同协议的网关通过分配不同的端口进行负载均衡。由于雷电大数据采集节点数较少,只需采用Keepalived+LVS这种轻量级的解决方案来保证LVS的高可用,即可避免出现单点故障。

#### 3.2 网关连接方式

雷电大数据采集系统网关主要处理雷电探测

终端上下行数据,指令下发。网关维持与终端的持久性链接,对接收到的终端数据进行简单的协议解析,并将接收到的数据持久化至消息中间件。

网关支持主动连接终端及被动连接两种方式,即可作为服务端等待终端连接也可作为客户端主动连接终端。这两种连接方式采用了不同的方案实现负载均衡,见图3。被动连接方式采用LVS实现负载均衡;主动连接方式网关通过Zookeeper监听服务数量,自动分配需要连接的雷电探测站,当新增网关或网关下线时,会重新分配雷电探测站。

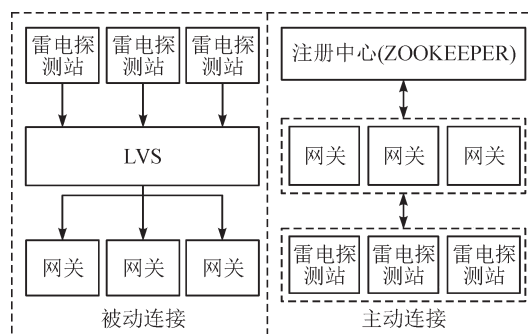


图3 网关连接方式

Fig. 3 Connection mode of gateway

系统接入服务网关基于高性能网络框架Netty开发,能够支持高并发的数十万计雷电探测设备接入,实现秒级下发指令,支持多种协议,因而通过配置可以实现不同类型雷电探测设备的接入。

#### 3.3 消息中间件

分布式系统构件之间通过传递消息可以解除相互之间的功能耦合,这样可以减轻子系统之间的依赖,使得各个子系统或者构件可以独立演进、维护或重用。

系统使用Kafka作为消息中间件。Kafka是一个消息发布订阅系统,见图4。Kafka中主要有3个角色:生产者、缓存代理和消费者。生产者向某个topic发布数据,订阅了这个topic的消费者就可以获取到相应的数据。

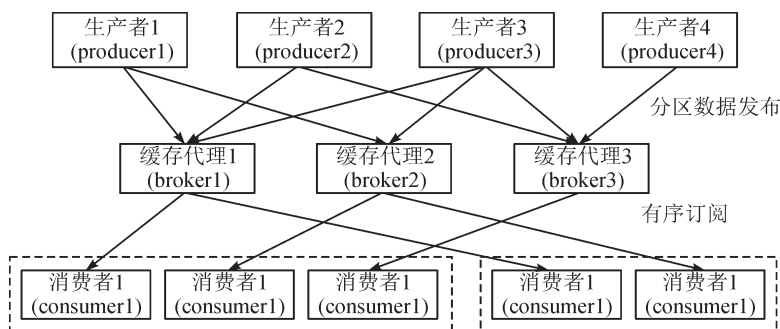


图4 Kafka架构图

Fig. 4 Architecture diagram of Kafka

当网络通道发生故障时,经网关过滤的原始数据缓存在kafka中,最大缓存容量取决于服务器硬件空间。当网络通道恢复时,Kafka中缓存的原始数据自动转发向下一环节,最大限度地保证了原始数据的完整。

### 3.4 实时计算框架

雷电大数据处理对计算的时效性要求很高,要求计算能在非常短的时延内完成。目前,实时计算比较优秀的框架主要有两个:Storm<sup>[19]</sup>和Spark Streaming。两者都可以提供很好的可扩展性和容错性,区别主要在于Storm处理的是每次传入的一个消息,而Spark Streaming是处理某个时间窗口内的所有消息,因此Storm对消息的处理可以达到毫秒级的延迟,而Spark Streaming则有秒级的延迟。依据雷电大数据采集对时效性的要求,选择Storm作为实时处理框架。

Storm架构图见图5。其中存在两类节点:主控节点和工作节点。主控节点上主要运行Nimbus,其主要职责是分发计算代码、在机器间分配计算任务以及故障检测等管理功能。ZooKeeper用来协调Nimbus和Supervisor。集群中的每台工作服务器上运行supervisor,其监听Nimbus分配给自己的任务,并根据要求启动或者停止相关的计算任务worker。

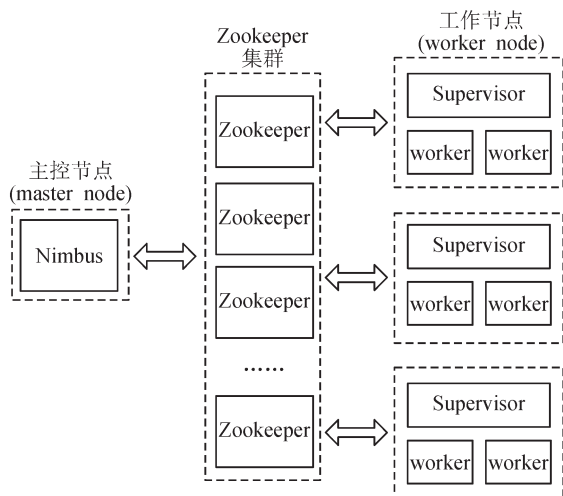


图5 Storm集群架构

Fig. 5 Architecture diagram of Storm cluster

### 3.5 分布式缓存

Memcached是分布式内存对象缓存系统,Redis是基于内存的键值存储系统<sup>[20]</sup>,两者性能都很出色,而后者支持更多样的数据类型,在内存空间上可以突破物理内存的限制。前者不支持数据持久化,断电或重启后数据消失,后者支持数据持久化和数据恢复,允许单点故障。

由于雷电数据不可再生,雷电大数据采集系统

选择使用Redis来存储实时计算解析后的终端雷电信息和状态信息。解析好的数据先缓存在redis中,再同步到关系型数据库中,从而在雷电高并发时缓解关系型数据库的读写压力,保证数据存储的稳定不丢失。

### 3.6 监控系统

Nagios是一个监视系统运行状态和网络信息的监视系统,能监视所指定的本地或远程主机以及服务,同时提供异常通知功能等。Ganglia是一个针对高性能分布式系统所设计的可扩展监控系统,它允许用户能够远程监控系统的实时或历史统计数据,包括:CPU负载均衡、网络利用率等。

电网雷电大数据采集系统采用Nagios + Ganglia对集群运行状态的性能指标、服务指标进行监控、故障报警等。

## 4 电网雷电大数据采集系统

基于分布式集群关键技术,研发了电网雷电大数据采集系统。该系统具有一站式数据接入、全方位接入管理、动态扩容机制、智能化运维等特性。

### 4.1 系统架构

电网雷电大数据采集系统架构由设备层、数据接入、数据计算、数据存储、数据交换接口、监控告警以及Web管理7个部分组成,见图6。

设备层包含通用服务器和交换设备;数据接入包括负载均衡(LVS/F5)、网关(GW)、消息中间件(Kafka)等模块;数据转发提供了数据路由服务,可将接受到的终端数据以及监控数据按照配置的路由规则转发给其他平台,同时也可接受其他平台转发的数据;WEB管理提供了终端管理、网关管理、实时计算任务管理、用户管理、数据交换等功能,实时计算用于实时解析终端原始报文、分析终端状态、报警信息等;基于数据接入,用户可从消息中间件订阅数据,实现数据分析、数据存储等功能,数据导入服务将用户关注的导入到存储中,支持Redis、Rdbms、Hadoop、Hbase等多种存储方式;数据交换接口实现了终端指令下发、原始报文查询、实时数据订阅、终端注册、注销等功能;平台监控采用Nagios+Ganglia,可监控每个节点的工作状态,并实现了系统、软件、业务以及硬件层面的全面监控,支持邮件、短信等多种告警通知方式。

### 4.2 系统数据流原理

系统的数据流图见图7。其中流程可描述为:

1)终端通过LVS负载均衡与网关建立长连接。

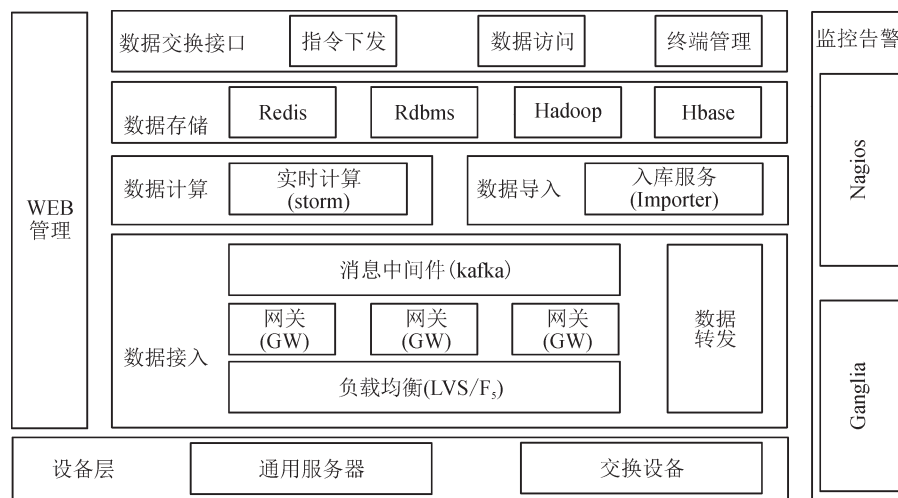


图6 系统架构图

Fig. 6 System architecture diagram

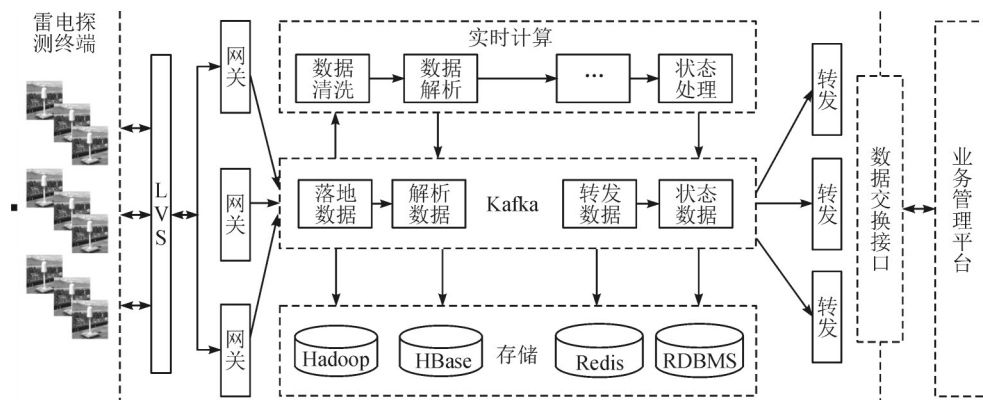


图7 系统数据流图

Fig. 7 System data flow diagram

2)网关通过对数据报文进行解码后封装成平台通用数据格式并写入Kafka。

3)实时计算模块通过订阅Kafka中的原始报文数据Topic进行数据解析,解析后的数据再次写入Kafka。

4)后续模块从Kafka订阅终端原始数据或解析后数据,实现数据存储或离线分析等功能。

5)转发服务从相应的Kafka Topic订阅数据并转发至其他平台。

6)业务管理平台可通过数据交换接口访问雷电大数据采集接入模块。

#### 4.3 业务管理平台

为实现对雷电大数据接入的可视化管理,设计开发用户业务Web管理平台,集中监控雷电探测站设备在线状态和数据采集服务的实时数据传输状态,使具备接入管理、实时计算、数据交换、集群管理、存储管理等功能,同时实现对集群节点和服务的监控及异常告警。其功能框见图8。

1)接入管理。该模块实现接入设备全方位管

理,具有分组管理、中心站管理、探测站类型、探测站管理、网关管理等功能。

2)实时计算。用户自定义快捷完成实时计算任务的创建、运行,注册探测站发送的数据完成原始数据的接收、解析。

3)数据交换。实现数据交换接口Redis数据查询功能和指令管理功能。提供Redis内存数据库中对对象记录查询的功能;进行指令下发,并实时展现指令下发后的状态和响应结果。

4)集群管理。实现系统提供服务的查看和管理、所有节点的信息查看功能。对系统所有提供的服务状态、节点信息、子服务信息、启停操作进行管理。对系统集群中所有节点的角色信息:主机名、IP地址、角色、CPU核数、磁盘空间、内存信息提供查看的功能。

5)存储管理。实现系统业务流转流程中对象的存储管理功能。在创建探测站类型、实时计算任务过程中,可以选择某些处理后的数据入Kafka,此时需要配置相应的Kafka Topic,默认情况下这些Topic

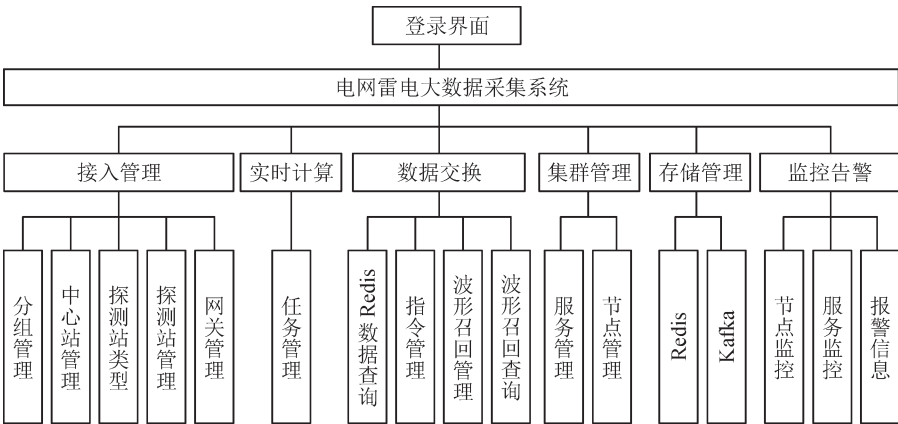


图8 业务管理平台功能框图

Fig. 8 Functional block diagram of operation control platform

的数据会存储在 Kafka 并存入 HDFS 中,用户可以根据需要让其存入 RDBMS(MySql/Oracle/SqlServer)与 BigTable。

6)监控告警。实现集群运行状态的性能监控、服务指标监控、故障报警功能。对系统涉及的整个集群、各个节点的负载、CPU、内存、网络性能指标的监控;对系统集群使用的关系型数据库、内存数据库 Redis 关键性能指标进行监控;对系统集群所有节点运行的监控指标进行监控告警,一旦有监控告警产生,以邮件或短信或其他自定义系统通知的方式提示用户。

5 应用实例

以该系统在某一省域电网的试点应用作为实例,验证系统的运行效果。该省域内现有雷电监测

低频雷电探测仪 24 个,其中模拟站 3 个,数字站 21 个;雷电预警传感站地面电场仪 6 个,每秒共产生雷电数据约 0.85 MB。在 2018 年 4 月底,低频雷电探测仪正常在线运行 18 个,在线率为 75%;雷电预警传感站正常在线运行 5 个,在线率 83.33%。现有低频雷电探测站与雷电预警传感站数据采集软件为分别独立开发,平台不统一,访问地址比较分散。引入雷电大数据采集系统后运行效果良好。

5.1 探测站信息录入

在接入管理功能模块中,依次新增分组、添加中心站;选择所属分组,建立探测站类型;选择所属中心站、探测站类型,录入探测站的详细信息,包含探测站 ID、探测站名称、探测站 IP、通讯端口号、连接方式、存储方式。录入完成后的探测站管理列表见图 9。图 8 中的列表可实时显示探测站的连接状态。

探测站管理(剩余可注册终端数 10)

| <div>新增 导入 修改 删除 导入模板下载</div> <div>选择探测站类型 探测站ID</div> |       |              |         |    |             |       |                  |      |    |
|--------------------------------------------------------|-------|--------------|---------|----|-------------|-------|------------------|------|----|
| 探测站ID                                                  | 探测站名称 | 探测站类型        | 所属中心站ID | 状态 | 探测站IP       | 探测站端口 | 连接方式             | 存储方式 | 操作 |
| 518                                                    | 探测站1  | nari_digital | 32      | 在线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 331                                                    | 探测站1  | nari_digital | 32      | 离线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 520                                                    | 探测站1  | nari_digital | 32      | 在线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 330                                                    | 探测站1  | nari_digital | 32      | 在线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 335                                                    | 探测站1  | nari_digital | 32      | 在线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 195                                                    | 探测站1  | nari_digital | 32      | 离线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 194                                                    | 探测站1  | nari_digital | 32      | 离线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 32                                                     | 探测站1  | nari_digital | 32      | 在线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 121                                                    | 探测站1  | nari_digital | 32      | 离线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |
| 120                                                    | 探测站1  | nari_digital | 32      | 在线 | xx.xx.xx.xx | 30001 | 探测站主动连接网关 落本地并转发 |      |    |

从1条到10条, 总共24长 (filtered from 10 total entries)

上一页 1 2 3 下一页

图9 雷电探测站信息列表

Fig. 9 List of lightning detection station information

### 5.2 实时计算任务添加

创建实时计算任务,实现原始接入数据的解析。分别创建雷电探测模拟站、数字站和雷电预警传感站实时计算任务,选择对应的sdk协议包,可对计算任务可进行启停操作,见图10。

### 5.3 数据实时存储

选择将解析后的数据存入数据库,数据库类型支持 MySQL、Oracle、SqlServer、BigTable,依次输入 IP 地址、端口信息、库名或SID信息、用户名、密码、

表名,见图11。在右侧设置好数据库列名对应的key值。

### 5.4 服务节点管理

当系统需要维护时,可对系统提供的服务节点进行启停操作管理,见图12。对于服务的启停操作需要严格按照顺序来执行:(Storm→Zookeeper | Kafka→Zookeeper),即停止操作需要先停止 Storm 和 Kafka,最后停止 Zookeeper,启动的时候顺序相反,否则系统会给出错误信息提示。

| 实时计算任务管理     |     |                     |          |         |
|--------------|-----|---------------------|----------|---------|
| + 创建任务       |     |                     |          |         |
| 任务名称         | 状态  | 创建时间                | 探测站类型    | 操作      |
| nari_old     | 运行中 | 2018-02-06 17:28:17 | nari_old | 操作> 查看流 |
| nari_predict | 运行中 | 2018-02-06 17:58:15 | nari_old | 操作> 查看流 |
| nari_gp      | 运行中 | 2018-02-06 17:58:45 | nari_old | 操作> 查看流 |
| nari_monitor | 运行中 | 2018-02-06 17:59:07 | nari_old | 操作> 查看流 |
| nari_digital | 运行中 | 2018-02-07 20:14:53 | nari_old | 操作> 查看流 |

图10 实时计算任务列表

Fig. 10 List of real-time computing task

nari\_digital\_state

入sqlServer ☒是 ☐否

类型 sqlServer

导入配置文件 ☐

IP地址 xx.xx.xx.xx

端口 xx33

库名或SID xxxxxxxx

用户名 xxx

密码 .....

表名 TDFV20\_Status

| 列名        | 列类型      | 对应key     | 操作 |
|-----------|----------|-----------|----|
| TDFTime   | datetime | tdfTime   | ✕  |
| HOSTID    | int      | hostId    | ✕  |
| TDFID     | int      | tdfId     | ✕  |
| TDFV20ID  | int      | tdfId     | ✕  |
| Longitude | int      | longitude | ✕  |
| Latitude  | int      | latitude  | ✕  |

图11 数据存储配置

Fig. 11 Data storage configuration

| 所属服务      | 进程名称             | IP          | 状态 | 操作  |
|-----------|------------------|-------------|----|-----|
| Storm     |                  |             | 正常 | 操作▼ |
|           | nimbus           | xx.xx.xx.xx | 正常 | 操作▼ |
|           | supervisor       | xx.xx.xx.xx | 正常 | 操作▼ |
| Zookeeper |                  |             | 正常 | 操作▼ |
|           | QuorumPeerMain   | xx.xx.xx.xx | 正常 | 操作▼ |
| Kafka     |                  |             | 正常 | 操作▼ |
|           | Kafka            | xx.xx.xx.xx | 正常 | 操作▼ |
|           | Kafka            | xx.xx.xx.xx | 正常 | 操作▼ |
| Importer  |                  |             | 正常 |     |
|           | BigTableImporter | xx.xx.xx.xx | 正常 | 操作▼ |
|           | RdbmsImporter    | xx.xx.xx.xx | 正常 | 操作▼ |
|           | BigTableImportec | xx.xx.xx.xx | 正常 | 操作▼ |
| Forward   |                  |             | 正常 | 操作▼ |
|           | Forward          | xx.xx.xx.xx | 正常 | 操作▼ |
| Schedule  |                  |             | 正常 | 操作▼ |
|           | TSchedule        | xx.xx.xx.xx | 正常 | 操作▼ |
|           | TSchedule        | xx.xx.xx.xx | 正常 | 操作▼ |

图12 服务节点管理界面

Fig. 12 Service node management interface

5.5 应用效果

雷电大数据采集系统试点应用以来,一直保持稳定运行。相对于传统的数据采集方案,由于平台统一,对雷电监测预警各类型探测设备的接入管理方便快捷。2018年5月,系统在雷电探测站未损坏情况下连接率接近100%,数据存储丢失率几乎为0,其中连接率为正常连接时间与总时间比率平均值,见表1。

表1 系统性能对比

Table 1 System performance comparison

| 性能指标                            | 原系统       | 文中系统      |
|---------------------------------|-----------|-----------|
| 最大并发数                           | 500       | 百万级       |
| 连接率/%                           | 95.89     | 99.66     |
| 数据丢失率/%                         | 2.00~4.00 | 0.02~0.05 |
| 数据处理速率/<br>(包·秒 <sup>-1</sup> ) | 200       | 50 000    |

当系统监控服务器临近发生异常或已发生异常时,系统管理界面及时提示用户,并发出对应的邮件通知,见图13。运维人员根据异常提示及时定位出系统问题出处,在异常发生前或刚发生时修复问题,提高了运维响应速度,多次及时保证了雷电系统的安全稳定运行。

| 主机名  | 状态 | 尝试次数 | 上次检查时间              | 检查命令                                   |
|------|----|------|---------------------|----------------------------------------|
| 服务器1 | 警告 | 4/4  | 2018-05-09 15:20:11 | check_nt!USEDIDKSPACE!-l d -w 80 -c 90 |
| 服务器1 | 异常 | 4/4  | 2018-05-09 15:18:06 | check_nt!MEMUSE!-w 80 -c 90            |
| 服务器2 | 异常 | 4/4  | 2018-05-09 15:18:45 | check_nt!MEMUSE!-w 80 -c 90            |
| 服务器3 | 警告 | 4/4  | 2018-05-09 15:20:42 | check_nt!USEDIDKSPACE!-l c -w 80 -c 90 |
| 服务器3 | 警告 | 4/4  | 2018-05-09 15:20:22 | check_nt!MEMUSE!-w 80 -c 90            |
| 服务器4 | 警告 | 1/1  | 2018-05-09 08:02:40 | check_stat                             |

图13 监控告警通知

Fig. 13 Surveillance alarm notice

参考文献:

[1] 陶汉涛,张 磊,陈 玥,等. 基于开源软件的电网雷电大数据平台技术研究与应用[J]. 智能电网,2017,5(8): 804-811.  
TAO Hantao,ZHANG Lei,CHEN Yue,et al. Research and application of lightning big data platform technology in power grid based on open source software[J]. Smart Grid, 2017,5(8):804-811.

[2] 谷山强,王 剑,冯万兴,等. 电网雷电监测数据统计与挖掘分析[J]. 高电压技术,2016,42(11):3383-3391.  
GU Shanqiang,WANG Jian,FENG Wanxing, et al. Statistical and mining analysis of lightning detection data in power grid[J]. High Voltage Engineering, 2016, 42(11): 3383-3391.

6 结论

文中阐述了雷电大数据的产生,说明了基于分布式集群的系统关键技术方案,研制了电网雷电大数据采集系统,并在江西电网进行应用实施,验证了系统具备以下优势:

- 1)基本克服了现有采集系统雷电数据短时集中高并发时可能引起的数据处理不及时、动态负载不均衡、断网时数据丢失等缺陷。
  - 2)实现了雷电数据一站式接入,从探测站终端数据监听读取到海量数据存储全自动化。
  - 3)实现了全程操控界面可视化,零代码管理,能对监控的服务及节点异常实时告警,变被动为主动,提前发现问题,使得系统的运行维护简单可控。
- 文中研制的电网雷电大数据采集系统具有通用性,其框架可广泛应用于电网数据采集领域;同时具有可扩展性,能便捷的增加不同协议的雷电探测设备,可为现有的雷电监测预警探测设备、研发中的高频及雷声探测设备甚至下一代雷电探测设备提供统一的数据接入平台。系统能有效保证雷电数据接入的实时性、完整性和准确性,为我国雷电监测预警网的安全稳定运行提供数据保障,具有很强的推广价值。

[3] 张启明,周自强,谷山强,等. 海量雷电监测数据云计算应用技术[J]. 电力系统自动化,2012,36 (24):58-63.  
ZHANG Qiming,ZHOU Ziqiang, GU Shanqiang, et al. Cloud computing application technology of massive lightning monitoring data[J]. Automation of Electric Power Systems,2012,36(24):58-63.

[4] 陶汉涛,冯万兴,卢恩泽,等. 输电线路跳闸雷电绕反击故障自动诊断系统的研发[J]. 高压电器,2016,52(5):34-41.  
TAO Hantao,FENG Wanxing, LU Enze et al. Research and development of automatic diagnosis system of lightning trip-out shielding failure and back striking fault for transmission line[J]. High Voltage Apparatus,2016,52 (5):34-41.

[5] 张东霞,苗 新,刘丽萍,等. 智能电网大数据技术发展研究[J]. 中国电机工程学报,2015,35(1):2-12.  
ZHANG Dongxia,MIAO Xin, LIU Liping, et al. Research

- on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-12.
- [6] 陈敬德, 盛戈皞, 吴继健, 等. 大数据技术在智能电网中的应用现状及展望[J]. 高压电器, 2018, 54(1): 35-43.  
CHEN Jingde, SHENG Gehao, WU Jijian, et al. Application status and prospect of big data technology in smart grid [J]. High Voltage Apparatus, 2018, 54(1): 35-43.
- [7] 季 艳, 鲁克文, 张英慧. 海量遥感数据分布式集群化存储技术研究[J]. 计算机科学与探索, 2017, 11(9): 1398-1404.  
JI Yan, LU Kewen, ZHANG Yinghui. Research on distributed clustering storage technology for massive remote sensing data[J]. Journal of Frontiers of Computer Science & Technology, 2017, 11(9): 1398-1404.
- [8] 郝春亮, 沈 捷, 张 珩, 等. 大数据背景下集群调度结构与研究进展[J]. 计算机研究与发展, 2018, 55(1): 53-70.  
HAO Chunliang, SHEN Jie, ZHANG Heng, et al. Structures and state-of-art research of cluster scheduling in big data background[J]. Journal of Computer Research and Development, 2018, 55(1): 53-70.
- [9] 杨际祥, 谭国真, 王荣生. 并行与分布式计算动态负载均衡策略综述[J]. 电子学报, 2010, 38(5): 1122-1130.  
YANG Jixiang, TAN Guozhen, WANG Rongsheng. A survey of dynamic load balancing strategies for parallel and distributed computing[J]. Acta Electronica Sinica, 2010, 38(5): 1122-1130.
- [10] 谷山强, 陈家宏, 陈维江, 等. 架空输电线路雷击闪络预警方法[J]. 高电压技术, 2013, 39(2): 423-429.  
GU Shanqiang, CHEN Jiahong, CHEN Weijiang, et al. Method for lightning flashover warning of overhead transmission lines[J]. High Voltage Engineering, 2013, 39(2): 423-429.
- [11] 陈家宏, 张 勤, 冯万兴, 等. 中国电网雷电定位系统与雷电监测网[J]. 高电压技术, 2008, 34(3): 425-431.  
CHEN Jiahong, ZHANG Qin, FENG Wanxing, et al. Lightning location system and lightning detection network of china power grid[J]. High Voltage Engineering, 2008, 34(3): 425-431.
- [12] 吴 健, 陈毅芬, 曾智聪. 利用地面电场仪与闪电定位资料进行短时雷电预警的方法[J]. 气象与环境科学, 2009, 32(1): 47-50.  
WU Jian, CHEN Yifen, ZENG Zhicong. Method of making a current pre-thundering report is to use the files of the ground electric field instrument and locating the lightning [J]. Meteorological and Environmental Sciences, 2009, 32(1): 47-50.
- [13] 陈家宏, 赵 淳, 谷山强, 等. 我国电网雷电监测与防护技术现状及发展趋势[J]. 高电压技术, 2016, 42(11): 3361-3375.  
CHEN Jiahong, ZHAO Chun, GU Shanqiang, et al. Present status and development trend of lightning detection and protection technology of power grid in China[J]. High Voltage Engineering, 2016, 42(11): 3361-3375.
- [14] 陶汉涛, 谷山强, 王海涛, 等. 输电线路雷电预警系统中心站的设计与研发[J]. 高压电器, 2016, 52(10): 61-68.  
TAO Hantao, GU Shanqiang, WANG Haitao, et al. Design and development of lightning warning system's central station for transmission line[J]. High Voltage Apparatus, 2016, 52(10): 61-68.
- [15] 杨铁军, 徐和飞, 黄 琳. 一种基于MINA框架的数据通信平台设计[J]. 微计算机信息, 2009, 25(11): 22-40.  
YANG Tiejun, XU Hefei, HUANG Lin. Design of data communication platform based on MINA[J]. Control and Automation, 2009, 25(11): 22-40.
- [16] 陶汉涛, 冯万兴, 谷山强, 等. 基于云计算的雷电监测预警与防护平台构建[J]. 高电压技术, 2017, 43(11): 3784-3791.  
TAO Hantao, FENG Wanxing, GU Shanqiang, et al. Construction of lightning monitoring, warning and protection platform based on cloud computing[J]. High Voltage Engineering, 2017, 43(11): 3784-3791.
- [17] 秦 刘, 兰巨龙, 杨 帅, 等. 动态反馈负载均衡在LVS集群中的设计与实现[J]. 电子技术应用, 2007, 33(9): 116-119.  
QIN Liu, LAN Julong, YANG Shuai, et al. Design and implementation of dynamic load balancing in LVS[J]. Application of Electronic Technique, 2007, 33(9): 116-119.
- [18] 马双良, 张英敏, 宋丽君. 基于LVS和计算任务的实时集群负载均衡方法[J]. 计算机工程与设计, 2007, 28(20): 4934-4937.  
MA Shuangliang, ZHANG Yingmin, SONG Lijun. Load balancing method for real-time cluster based on LVS and handling tasks[J]. Computer Engineering and Design, 2007, 28(20): 4934-4937.
- [19] 鲁 亮, 于 炯, 卞 琛, 等. 大数据流式计算框架Storm的任务迁移策略[J]. 计算机研究与发展, 2018, 55(1): 71-92.  
LU Liang, YU Jiong, BIAN Chen, et al. A task migration strategy in big data stream computing with storm[J]. Journal of Computer Research and Development, 2018, 55(1): 71-92.
- [20] 彭灿华. Redis在高速缓存系统中的序列化算法研究[J]. 现代电子技术, 2017, 40(22): 122-124.  
PENG Canhua. Research on serialization algorithm of Redis in cache systems[J]. Modern Electronics Technique, 2017, 40(22): 122-124.
- 朱 晔(1976—), 男, 高级工程师, 主要从事输电线路状态监测技术、电力系统自动化软件研发。
- 姜志博(1987—), 男, 硕士, 工程师, 主要从事电力信息与通信技术、雷电监测预警及防护技术的研究(通讯作者)。