

智慧供热大数据监测平台研究及应用

李恩洲,况立群*,张元,韩燮,熊风光

(中北大学大数据学院,山西太原 030051)

摘要:针对供热环境在城市间一体化程度的提高而引发的设备种类繁多、采集信息量大、数据类型多样化、处理和分析困难以及各设备数据信息展示不直观的问题,研发并实现了一种智慧供热大数据监测平台,并在山西省供热体系中进行应用。平台以云计算为基础架构,采用基于 React 的前端技术,将开发模块组件化,并通过构建虚拟 DOM(虚拟文档对象)机制解决了因复杂和频繁的 DOM 操作引发的性能瓶颈问题。针对供热设备产生数据的结构特点,研发基于 kafka+Spark Streaming 的数据清洗方案,解决了因数据量大、种类多样化而产生的数据处理效率、存储设备压力以及实时数据传输问题。此外,平台还提供了定制化的可视化界面,并且通过数据分批处理机制,解决了实时数据和大数据量展示过程中响应慢的问题。

关键词:智慧供热;监测;React;可视化;B/S;kafka;Spark Streaming

中图分类号:TP311.5

文献标识码:A

文章编号:1673-629X(2021)11-0176-07

doi:10.3969/j.issn.1673-629X.2021.11.029

Research and Application of Big Data Monitoring Platform for Intelligent Heating

LI En-zhou, KUANG Li-qun*, ZHANG Yuan, HAN Xie, XIONG Feng-guang

(School of Big Data, North University of China, Taiyuan 030051, China)

Abstract: Aiming at the problems of various types of equipment, large amount of collected information, diversification of data types, difficulties in processing and analysis, and unintuitive display of data information of various equipment caused by the increase in the degree of integration of thermal companies in various regions, we research and implement a smart heating big data monitoring platform, which is applied in the heating system of Shanxi Province. Based on cloud computing and React front-end technology, the platform componentizes development modules and solves performance bottlenecks caused by complex and frequent DOM operations by building a virtual DOM (virtual document object) mechanism. In view of the structural characteristics of data generated by heating equipment, a data cleaning solution based on kafka+Spark Streaming is developed to solve the problems of data processing efficiency, storage equipment pressure and real-time data transmission caused by the large amount of data and the diversification of types. In addition, the platform also provides a customized visualization interface and solves the problem of slow response during real-time data and large data presentation through the data batch processing mechanism.

Key words: intelligent heating; monitor; React; visualization; B/S; kafka; Spark Streaming

0 引言

近年来,国内外针对供热系统的研究主要集中在供热设备以及故障检测方面^[1-3],其中在智能故障诊断系统的研究中,对换热器的诊断已经形成了良好的生态系统^[4]。文献[5]针对集中供热系统中多热源调度优化模型进行了参数的调节,使得整个运行环境从能耗和污染物排放量方面得到了最优解。对于智慧供热解决方案的研究,主要面向特定区域、特定的供热环

境,实现了从热源、管网、场站,到户内的数据互联互通与共享。

而对于跨区域及城市的供热管理体系,目前尚未形成统一的标准。对于省市一体化的供热监测还未形成成型的系统。并且随着地域的扩大,使用单一的数据库服务已经不能满足供热设备集群的实时数据以及元数据的处理,由设备集群产生的高并发数据势必会使数据库服务器产生性能问题。文中旨在研究一种智

收稿日期:2020-11-23

修回日期:2021-03-23

基金项目:国家重点研发计划(2018YFB2101504)

作者简介:李恩洲(1994-),男,硕士研究生,研究方向为数据可视化、数据挖掘;通讯作者:况立群(1976-),男,副教授,CCF会员(48959M),研究方向为人工智能与计算机视觉。

慧供热大数据平台,通过构建 kafka+Spark Streaming 处理机制解决海量实时数据和离线数据处理,对于供热环境的外部影响因素也可以依据该平台进行相关分析。同时,定制化的将供热系统的重要参数布局在可视化页面中,便于实时查看。另外,对相关参数实施预警处理,使得通过互联网对供热系统的故障诊断更加方便快捷。

1 平台概述

平台为整个供热环境提供了一系列的配套服务,包括整体环境如天气、供热面积、能耗、换热站数据分

析、数据展示等,这些服务可以为供热管理人员提供便捷的管理,使各项工作化繁为简。系统通过对供热设备中已有的数据采集设备与供热设备相关的配套服务进行数据交互,将供热设备中产生的数据进行清洗、分析和计算,一方面将处理结果传输至相应的应用,另一方面将数据持久化,存储至数据库中。数据以 json 数据形式进行传递^[6],并将数据进行格式化处理,从而解决数据可视化中的数据统一问题,最后传入到客户端。管理人员可以在系统中监控和观察设备,以方便工作人员进行相应的处理,具体业务框架如图 1 所示。

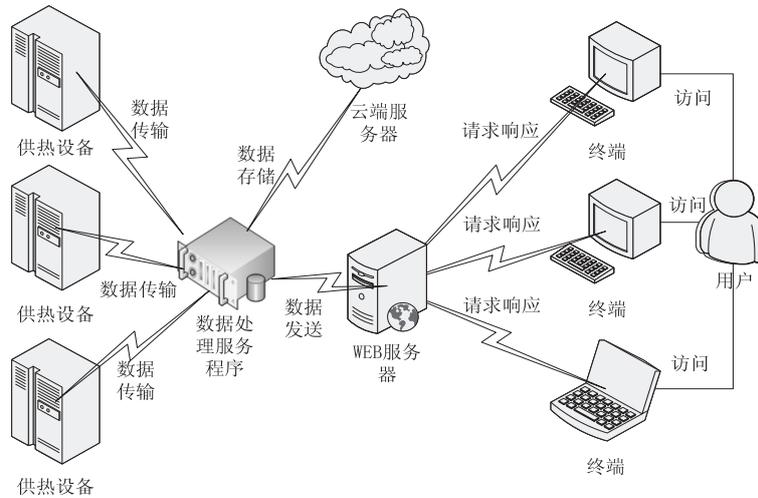


图 1 平台整体业务框架

2 平台架构设计

平台架构以云计算为基础,云计算通过 Internet 提供快速创新,弹性资源和规模经济,对于云服务,通常按照使用量支付费用,从而帮助降低运维成本,使基础

设施更有效的运行,并能根据业务需求的变化调整对服务的使用^[7],云计算服务提供了基本的四大类服务:基础结构即服务(IaaS)、平台即服务(PaaS)、无服务器计算和软件即服务(SaaS),具体平台架构如图 2 所示。



图 2 云平台系统结构

2.1 感知传输层

感知层即数据采集层,是热源、热网、换热站、末端用户等供热环节中的各类监控系统和管理服务的整合,其中包括 PLC 控制系统、数据采集系统、数据交互系统、用户交互系统、地理信息系统、设备监控管理系统、视频监控系统等业务模块,它将物理、交互、音视频等各类型信息按照统一标准化对数据进行清洗、过滤、转换,通过传输层将数据传输到中心平台。

2.2 基础设施层

基础设施层(IaaS)是系统业务的底部支撑载体,内部包括物理资源层、资源虚拟化层和云服务管理层。

(1)物理资源层:指的是计算服务集群,包括热力能源 web 服务器、后端服务器、数据库服务器。

(2)资源虚拟化层:是云计算的根本技术支撑,存储方面则主要使用分布式存储架构共享存储集群,用来针对不同的存储数据,例如换热站供温供压等参数数据的分类存储。

(3)云服务管理层:为 IaaS 架构中最为关键的一层,通过云服务管理可以对供热设备进行开关控制、权限设置、故障恢复等管理。

2.3 平台服务层

PaaS 提供了应用程序和开发环境,将各种业务能力进行整合,具体可以归为热力能源 web 服务器、二次能源系统开放平台、热网运行监控系统开放平台、设备管理系统开放平台,向下根据业务能力需要测算基础服务能力;通过 IaaS 提供的 API(application programming interface,应用程序接口)调用硬件资源,向上提供业务调度中心服务,实时监控平台的各种资源,并将这些资源通过 API 开放给 SaaS 用户。

2.4 软件服务层

软件服务层(SaaS)在 PaaS 层之上,它将智慧供热系统下的子系统,以服务的形式发布在云平台之上,便于企业更好地规划系统模块结构,进行及时调整和快速部署。这些子系统包括热网监控系统、全网平衡分析系统、在线热网水力分析系统、热网能耗分析系统与供热地理信息系统、热用户室内测温、热计量系统、气象数据分析系统。

3 平台的实现和应用实例

文中以山西省供热体系为例进行热力能源管理平台实例分析,针对供热系统中管理员对整个供热环境的态势感知和供热设备实时数据和离线数据处理困难的问题。并且考虑到管理人员需要进行相关的汇报工作,故选用开源 web 框架 React 作为数据可视化的实现基础,以 G2 数据可视化组件进行数据的多元化展示,具体方案如下。

3.1 需求分析

通过调研山西省各大热力公司的数据监控业务,总结出软件主要功能需求如下:

(1)用户管理:对使用平台的用户进行管理和授权。

(2)设备信息提取:对设备产生的相关数据进行提取清洗和分析,并将相关结果存储在数据库中。

(3)设备远程控制:从不同的维度出发,进行设备的远程控制,包括开关控制、参数设置、预警控制。

(4)热力系统重要参数的运行数据以及信息的可视化。

(5)展示相关换热站的数据信息。

(6)能耗相关数据信息。

(7)系统操作日志的记录。

3.2 功能设计

3.2.1 用户管理

对每一个供热地点,系统都会为其分配一个管理员账户,管理员账户不可删除,具有最高权限,可以管理其下面的所有账户,在管理员账户之下可以申请二级账户,方便工作人员的管理,同时对供热设备的故障处理信息进行记录。对于管理账户,可以进行所有设备数据操作,并生成日志记录。管理员账户可以新增或者删除权限较低的账户。

3.2.2 数据处理模块

关于供热设备的数据清洗和处理,文中采用 kafka+Spark Streaming 的方式进行实时数据流式处理,对于大量设备的原始数据进行高并发的清洗和转换,然后将数据通过统一结构化模型的 api 实时显示在 web 端,离线数据可通过对应的 api 进入到相关数据库中进行存储,具体流程如图 3 所示。

3.2.3 可视化模块管理

对于供热系统中概览数据展示设计为九宫格的形式,主要展示对于供热系统相关参数的分析,包括天气情况、热源实时数据、热源历史数据、供热地区的排行、告警信息、能源消耗占比、室温实时数据以及对总体环境的数据的展示。

(1)天气的展示主要以折线图的方式进行,展示从当天起前七天的最高温和最低温,具体的天气的分析会在单独的页面进行管理和查看。

(2)热源数据的展示分为实时数据和历史数据。实时数据主要运用对照的方式进行展示,例如供水温度与回水温度的对比。考虑到实时数据的特点,将折线图进行改进,加入了标准线以及实时数据的刷新,对于触及报警线及以上的数据,进行预警处理;对于热源历史数据的展示主要是选择从当天起前七天数据作为展示周期,对比数据分析从而了解供热设备系统的运

行状况,进行针对性的调整。

(3)数据榜单的设计主要以供热系统所管理的区域进行,按照换热站的供热情况,取较好和较差的供热表现的换热站进行数据展示,在获取数据的过程中进行排序以及数据处理。

(4)告警信息和室温实时数据的处理以列表的方式进行滚动轮播,考虑到供热系统管理区域的小区数量以及实时更新的特点,相比柱状图、饼图、雷达图等

图形化展示方式更有利于工作人员进行观看,可以展示各个区域的供热情况,减少了工作人员的操作,提高了便捷性。

(5)系统在板块划分中加入了总体实时数据的一个展示,展示的内容主要包括热源数量、换热站数量、天气信息、供热面积以及所有能耗的总数和设备负荷的相关数据,便于工作人员判断设备的运转情况。

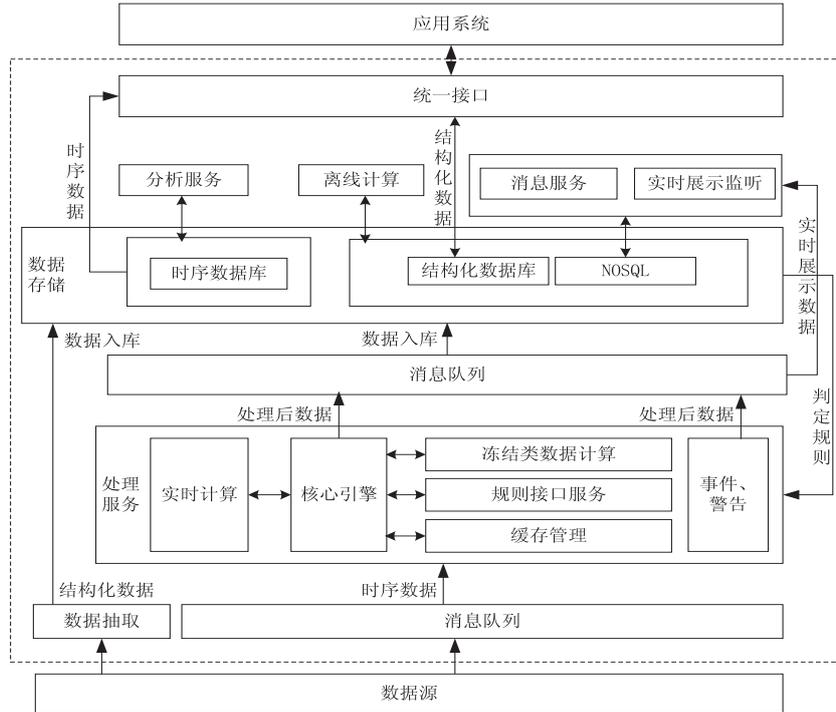


图 3 平台数据流程图

3.2.4 供热设备数据管理

对于换热站数据的管理包括报表数据、操作日志数据,对于报表主要展示的数据为换热站的基础数据,操作日志主要展示的数据为某一管理人员对换热站的相关操作并且记录了改动前后的变化。

3.2.5 能耗管理

能耗管理模块中有三个功能:综合报表、环比报表和能耗排行。综合报表主要显示的是所有换热站在某一时间段的热、电、水耗的相关数据,用户可以查询任意时间段内的数据;环比报表在显示能耗的相关数据的基础上加入了时间段的比对,可以比对任意两个时间段的换热站的能耗数据,用于更好地管理供热系统。同时两个报表提供了下载和打印功能,可以直接将查询的数据转换为 excel 表格。此外,该模块提供在线打印功能。

3.3 软件使用相关技术

3.3.1 基于 React 的前端技术

在 Web 开发中,需要将变化的数据实时反映到 UI (user interface) 上,这时就需要对 DOM (document

object model) 进行操作,而复杂且频繁的 DOM 操作通常是性能瓶颈产生的原因。因此,文中采用的 React 框架中引入了虚拟 DOM (virtual DOM) 的机制:在浏览器端用 Javascript 实现了一套 DOM API^[8]。基于 React 进行开发时,所有的 DOM 构造都是通过虚拟 DOM 进行,每当数据发生变化时,React 都会重新构建整个 DOM 树,然后将当前 DOM 树和上一次的 DOM 树进行对比,得到两者的不同,然后将需要变化的部分做浏览器 DOM 更新。

同时,为了使得前后端有效地进行交互,引入 Node.js,一个事件驱动 I/O 服务端 JavaScript 环境,基于 Google 的 V8 引擎,该框架善于处理高并发需求、单线程、拥有事件机制以及跨平台的特性,使得它在数据通信频繁的实时业务场景下产生了较大的优势^[9]。Node.js 的最重要的功能是事件驱动的,这意味着服务器仅在用户有事件时才响应。以 Node.js 充当浏览器的服务器,将数据转发到后端服务器,并通过路由的方式控制页面跳转的逻辑。

而 antdesignpro 这套企业级中后台前端开发框架

应用了 React 框架的相关技术并且在 antdesign 的设计规范和基础组件的基础上进行进一步封装^[10]。antdesignpro 结合 dva 框架,通过 reducers、effect 和 subscriptions 组织 model;基于 babel-plugin-dva-hmr 实现 components、routes 和 models 的 HMR (hot module replacement),既满足了 MVC(model-view-controller) 中的‘V’即视图部分^[11],同时又满足了前后端数据交互的部分,具体原理如图 4 所示。

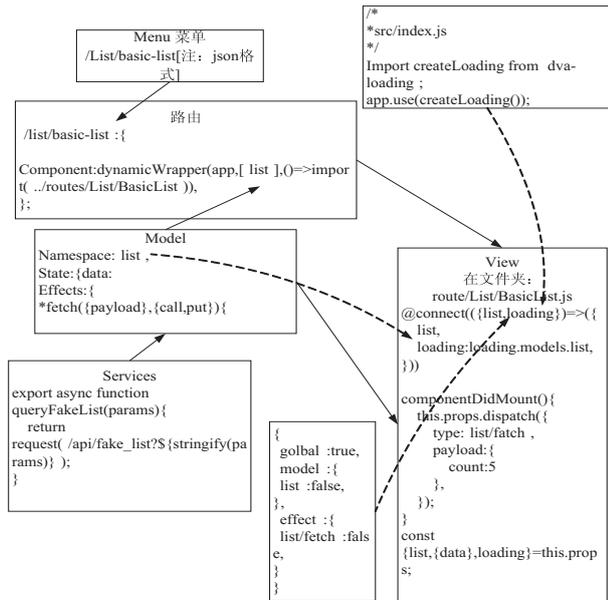


图 4 React 架构原理

平台将基于 React 的前端框架进行研发,优化内容主要有可视化组件封装、主题切换预留接口、table 标签二次封装。

对于可视化组件的封装,以 bizcharts 为主,将 bizcharts 图形的基本类型:折线图、柱状图、饼图、条形图封装成组件,通过引用相应的组件即可进行个性化的定制,并且提供了图形基本类型的样式,供研究人员参考。

对于主题切换,根据不同热力公司地区和行业特点开发了科技蓝和庄重灰等多种主题风格供热力能源服务商进行选择。此外,该模块提供统一接口,使得服务商可针对不同地区,不同环境进行定制化设计,以此增加平台的可复用性。

对于报表的展示,除了对于数据的显示,还封装了数据过滤、排序、文本类型、表格风格等功能,使得平台可以根据热力站的需求调用相应 API 进行相应的设置,从而更好地进行数据展示。

3.3.2 大数据实时可视化展现技术

对于供热系统中重要参数的实时大数据在可视化页面中的展示,文中基于 G2 的 React 封装图表组件库^[12] BizCharts 来实现。通过 BizCharts 中提供的 API: Guide 属性来绘制图表的辅助元素,达到图表定制化

设计,同时研发 dataSet 独立数据模块,为数据可视化场景提供状态驱动。文中研发的大数据可视化展示技术,首先将数据分批请求,通过限定请求时间范围的大小,将其分成若干份 Request,其次循环异步批次处理,后端将请求的数据以数据流形式发送,将得到的数据存储为 csv 文件,文件命名方式通过请求参数命名,最后在视图驱动程序中加载 csv 文件即可,同时,在请求前需进行 csv 文件扫描,查找是否存在查询的记录,如果存在,则直接加载文件,避免了多次请求服务器引起的带宽和并发。

3.3.3 驾驶舱自由布局

平台支持自定义概览主页面,即驾驶舱,通过该功能,用户可以在面板上自由的拖拽,修改页面中的组件,从而呈现出个性化的监控页面。React-Grid-Layout 是基于 React 的网格布局系统,支持响应式,在组件改变大小的时候,内部的图形和内容会随着组件的改变而进行响应式变化,并且支持断点,断点布局可由用户提供或自动生成,而且不需要 jQuery 支撑,其中有几个重要的函数用于布局中:

- (1) WidthProvider: 一个用于响应元素高度的高级函数,主要作用是绑定 onresize 事件,当组件改变时,用于重新计算元素的宽度。
- (2) GridItem: 根据 x、y 坐标计算出组件所在的位置。

具体流程如图 5 所示。

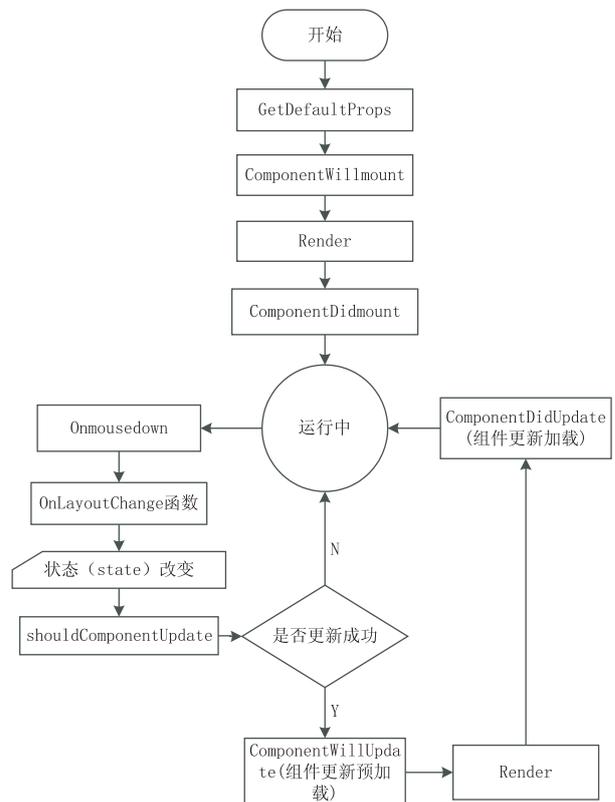


图 5 驾驶舱自定义实现流程

3.3.4 数据处理技术

从研究和商业水平的角度看,云一直是信息技术领域最热门的讨论主题之一。但是,对于云存储而言,没有主流的统一存储解决方案^[13],使得云端存储大量冗余数据,影响其性能。故采用 kafka+Spark Streaming 的设计方案,对供热设备元数据进行清洗和处理^[14],使得大量的数据无须存储在云端,仅需要把消息队列中的数据进行分布式存储,减小了对云存储的压力。

kafka 是一个高吞吐的分布式消息队列系统,特点是生产者消费者模式,先进先出(FIFO)保证顺序和数据不丢失,默认每隔 7 天清理数据。

其中主要分为两部分:producer 为消息生产者,Consumer 为消息消费者,在系统中,将 Spark Streaming 作为 Consumer。Broker 为 kafka 集群的服务程序,负责处理消息读写请求和存储消息。在 kafka cluster 这一层里,有多个 Broker。元数据信息存在 zookeeper 中,包括存储消费偏移量、topic 话题信息、partition 信息。

Spark Streaming 支持从多种数据源获取数据,其中一个核心为支持一个高层的抽象,称为离散流(Discretized Stream)也称作 DStream,代表连续的数据流。在内部,DStream 是由一系列 RDD 组成的。一批数据在 Spark 内核中对应一个 RDD 实例,因此,对应

流数据的 DStream 可以看成是一组 RDD,即一个 RDD 的序列。也就是说,在流数据分成一批一批后,生成一个先进先出的队列,Spark Engine 从该队列中依次取出批数据,并把批数据封装成一个 RDD,然后再进行处理,最后将结果进入统一 services 接口,将数据格式化后进行实时显示,或者将数据进行存储。确保数据的有效性和实时性。

3.3.5 SpringBoot

Spring 开源框架是基于 J2EE 框架的研发产品。其中 Spring MVC 框架(MVC 是模型视图控制器,通过将业务逻辑,数据和界面显示分离的方法来组织代码)被广泛用于各种 Web 产品的研发中。SpringBoot 项目是 Pivotal 团队开发的新 Spring 框架。其开发的目的是解决 Spring 开发过程中的复杂配置操作并最终实现零配置。其简化的应用程序开发和构建过程使开发人员摆脱了繁琐的配置文件,并投入了大量精力进行项目逻辑研究和开发。它致力于成为快速应用程序开发的领导者^[15],同时,其内嵌的 Tomcat 服务器,直接将项目打包成 jar 包,从而简化了项目的部署工作。

3.4 平台实例展示

系统的具体效果如图 6 所示。

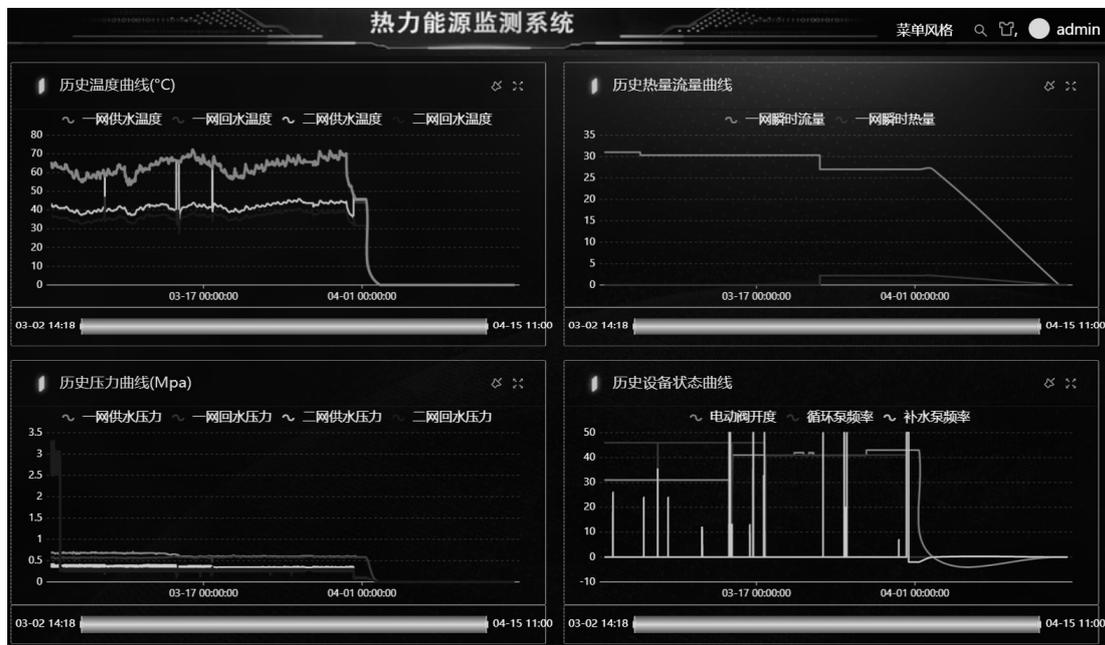


图 6 加时间进度条的可视化数据显示

考虑到供热系统的时间维度问题,故在平台中设计了历史数据曲线可视化模型,具体效果如图 7 所示,其中顶部进行换热站以及时间段的选择。为了避免时间跨度较大导致的卡顿现象,设计了时间间隔,以应对大数据量导致的数据传输问题。可视化组件以折线图为基础模型,在模型之上设计了时间进度条,通过拖拽

进度条可以显示该时间段内的任意段时间的数据,平台展示出整体供热系统的历史温度、历史热量流量、历史压力、历史设备状态,使得管理者可以更加直观地进行系统的分析与诊断。

平台提供了不同维度的设备监控,并且通过点击设备的组件,实现远程控制设备,具体设计如图 7

所示。

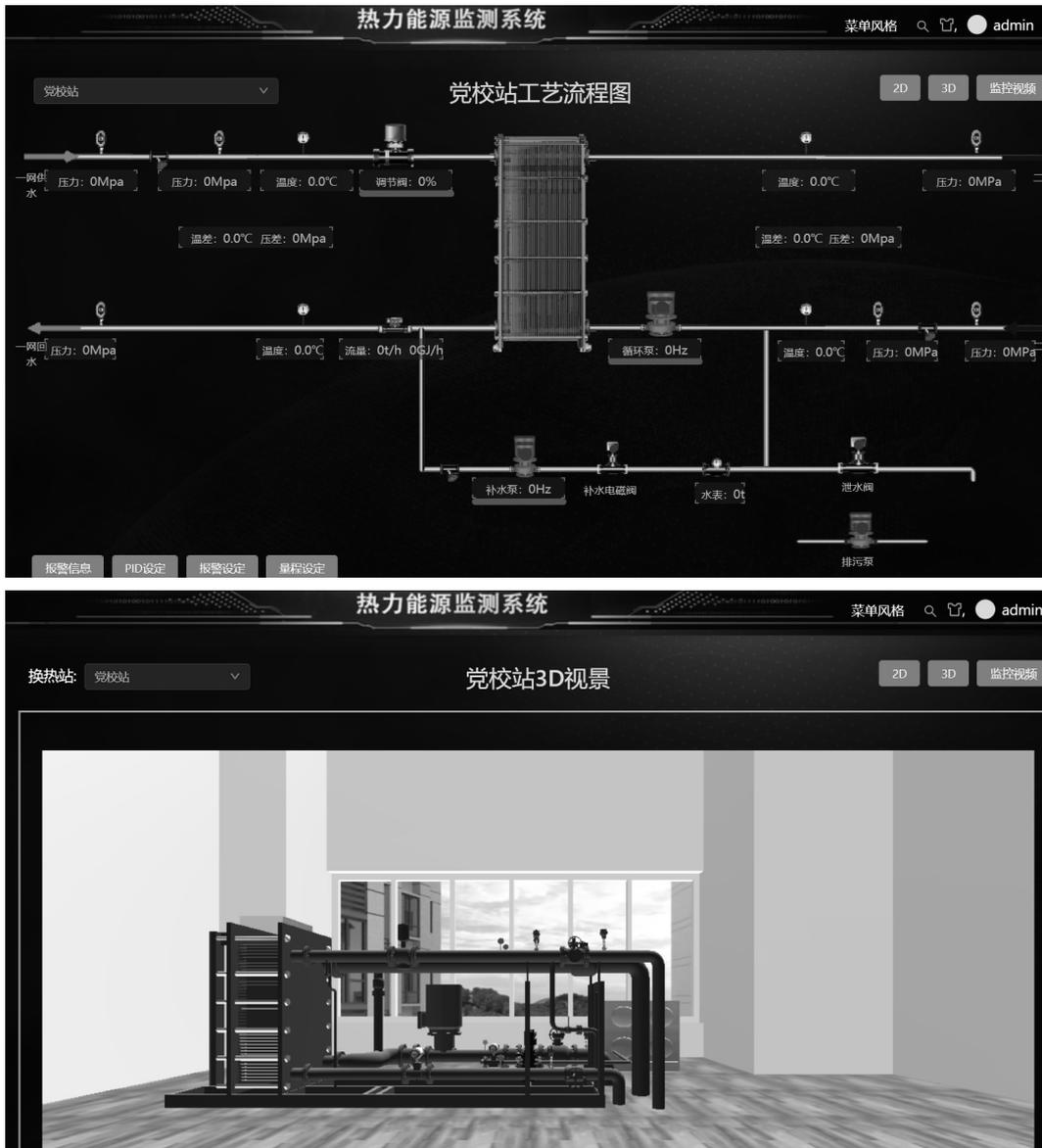


图 7 加时间进度条的可视化数据显示

经测试,系统无异常,并且对于页面的显示和用户交互方面与系统需求一致。

4 结束语

供热平台对供热系统工业自动化、信息化、智慧化发展,起到了一定的推动作用,整合现有资源打造智慧供热是供热产业发展的未来方向。文中基于 React 框架,实现了供热环境的数据处理、展示以及监测供热过程,并且针对供热系统的特殊性,研发出定制化的可视化组件,最后,将平台应用于山西省的供热系统中,验证了平台的可用性。

参考文献:

[1] 张伟,袁艳平,贺文晟,等.高原河谷地区地下水源热泵供暖系统设备选配及优化运行[J].建筑科学,2015,31

(12):135-140.
 [2] 李亚军,谭荣帅,梁文豪.热泵在热电联产集中供热系统中的应用及(火用)分析[J].华南理工大学学报:自然科学版,2017,45(10):8-15.
 [3] 戈志华,孙诗梦,万燕,等.大型汽轮机组高背压供热改造适用性分析[J].中国电机工程学报,2017,37(11):3216-3222.
 [4] 安艳平,田琦,程锁明.板式换热器智能故障诊断系统研究[J].中国科技论文,2015(11):1291-1294.
 [5] 时国华,杨林棣,张浩,等.集中供热系统多热源调度优化模型[J].热力发电,2020,49(3):68-75.
 [6] BAAZIZI M A, COLAZZO D, GHELLI G, et al. Parametric schema inference for massive JSON datasets [J]. VLDB Journal, 2019, 28(4):497-521.
 [7] COLOMBO-MENDOZA L O, VALENCIA-GARCÍA R, COLOMBO-PALACIOS R, et al. A knowledge-based multi-

(下转第 188 页)