

陈剑飞,史彩霞,王振国. 基于行业气象服务的基础数据支撑平台研发与应用[J]. 气象研究与应用,2022,43(2):117-121.
Chen Jianfei, Shi Caixia, Wang Zhenguo. Implementation and application of basic data support platform based on industry meteorological services[J]. Journal of Meteorological Research and Application, 2022, 43(2): 117-121.

基于行业气象服务的基础数据支撑平台研发与应用

陈剑飞, 史彩霞*, 王振国

(广西壮族自治区气象灾害防御技术中心, 南宁 530001)

摘要: 以行业气象服务为依托,采用 Java、WEB 等技术,建立了基础数据支撑平台,并对其构建流程和功能模块进行了详细介绍。本系统采用 B/S+C/S 混合架构,各层分别独立,以提高系统的安全性、稳定性和可维护性。通过基础数据支撑平台,快速完成对各业务服务系统和专线用户的数据支撑,为做好行业气象服务提供借鉴和参考。
关键词: 行业气象服务;大数据;“天擎”;可视化
中图分类号: P49 **文献标识码:** A **doi:** 10.19849/j.cnki.CN45-1356/P.2022.2.20

前言

气象与各行各业都密切相关。广西气象部门依托于全区海量的气象数据研制行业气象服务产品,通过各类气象服务系统和专线为众多部门提供气象服务保障。目前建成的电力专业服务、交通服务等各类服务系统^[1-6]在气象现代化和行业服务中发挥了重要作用,但随着业务发展、网络安全和业务规范不断调整的同时也带来了一些问题:一是各业务服务系统的气象业务数据、用户服务数据、地理信息数据等都是相对独立的,“信息孤岛问题”提高了运营与维护成本,未能形成统一高效的服务供给;二是各业务服务系统的信息共享与协同访问很难实现;三是分布在各个部门的业务数据没有整合和统一规划设计,不能全面高效地支撑气象服务应用的开展。因此,通过搭建基于行业气象服务的基础数据支撑平台,打造从数据采集解析推送的全流程可视化监控和自动告警,实现对各类数据的采集挖掘,气象和行业数据的融合与共享应用,为广西行业气象服务提供有力的业务支撑。

1 基础数据支撑平台总体结构

基础数据支撑平台整体框架如图 1,工作方式采用 B/S+C/S 混合架构,采用 JAVA、WEB 等技术实现气象信息采集、解析、推送、监控、业务服务系统管理等功能。网络设计具有良好的可扩展性。平台按功能划分为数据采集解析系统、大数据推送管理系统、数据监控平台和业务系统综合管理平台四部分。

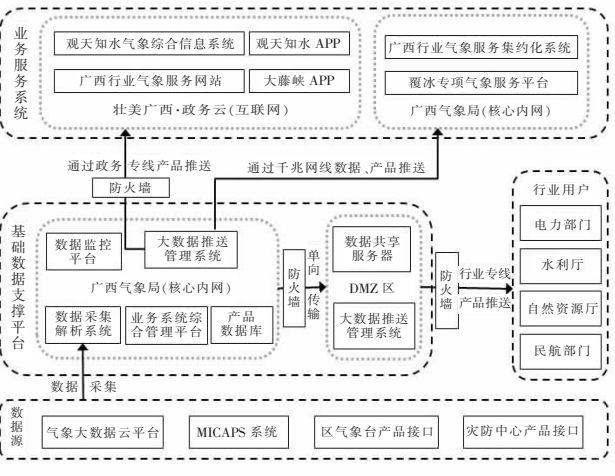


图 1 基础数据支撑平台整体框架图

收稿日期: 2022-05-08
基金项目: 2020 年广西区部合作项目“行业气象服务基础支撑平台建设”、广西气象科研重点项目(桂气科 2021Z04)
作者简介: 陈剑飞(1975—),男,高级工程师,主要从事专业气象技术服务及管理工作。E-mail:497933494@qq.com
* 通讯作者: 史彩霞(1975—),女,高级工程师,主要从事专业气象服务系统研发工作。E-mail:14557000@qq.com

(1)基于现有业务的布局,目前的网络分为广西气象局核心内网区、广西气象局 DMZ 区、广西大数据发展局壮美广西·政务云互联网区、行业专线对应的行业用户区共四部分。在核心内网与 DMZ 区、互联网区分别架设防火墙,配置数据只能从高安全区(核心内网)向低安全区(DMZ 区、互联网区)流动的安全策略,在 DMZ 区与行业用户区架设防火墙并根据业务需要配置相应安全策略,提高网络安全。

(2)大数据推送管理系统(内网)通过政务专线将服务产品推送到对应的政务云数据库服务器上,实现互联网服务系统的数据支撑。

(3)大数据推送管理系统(内网)通过千兆网线将服务产品推送到预报员使用的制作平台数据库服务器上,实现内网业务系统的数据支撑。

(4)大数据推送管理系统(内网)通过防火墙,单向传输产品到 DMZ 区数据共享服务器。

(5)大数据推送管理系统(DMZ 区)通过防火墙和行业专线,基于 SFTP 加密传输协议,将产品推送到各行业用户服务器,从而保证了数据安全。

(6)通过搭建服务器集群,分别部署分布式的采集服务、解析服务、推送服务,进行协同工作,从而保证某台服务器发生故障时其他服务器能继续处理任务,实现系统的稳定运行。

2 采用的关键技术

2.1 Kafka 消息中间件系统

Kafka^[7]是一种发布/订阅(Pub/Sub)模式^[8]的消息中间件系统^[9],以集群方式部署,主要应用于大数据实时处理领域,连接系统中的数据上下游,处理活跃的流式数据。Kafka 可降低系统组网复杂度和编程复杂度,数据上下游的应用程序只需要与 Kafka 提供的 API 对接即可,保证系统的高吞吐、高可用以及高容错性^[10]。Kafka 具有同时为发布和订阅提供高吞吐量,进行集群间的消息同步,并将消息持久化到本地日志等。Kafka 是分布式集群系统,含有多个服务器实例 broker,每个主题 topic 会有多个分区并均匀地分配到整个集群中,当生产者向对应主题传递消息,消息通过负载均衡机制传递到不同的分区以减轻单个服务器实例的压力^[11]。

2.2 Zookeeper 分布式协调服务

ZooKeeper 是一个开放源码的分布式应用程序协调服务,它包含一个简单的原语集,分布式应用程序可以基于它实现同步服务,配置维护和命名服务

等,具有类似文件系统、以节点为基础保存数据的结构,以及当目录节点发生变化(数据改变、被删除、子目录节点增加或删除)时的通知机制。Zookeeper 可提供一种可靠的、可扩展的、分布式的、可配置的协调机制来统一系统的状态^[12-13]。

2.3 Redis 缓存技术

Redis^[14](Remote Dictionary Server),即远程字典服务,是一个开源的使用 ANSI C 语言编写、支持网络、可基于内存亦可持久化的日志型、Key-Value 数据库,并提供多种语言的 API。它支持多种数据类型的存储和丰富的操作,其中大多都支持原子操作;将数据缓存在内存中,从而保证数据存取的效率;提供对持久化的支持,可定期将更新的数据异步写入磁盘,同时不影响继续提供服务;Redis 集群采用主从形式,确保当主节点失效后可以将一个从节点转变为主节点,从而有效预防单点故障和提高负载能力,以此确保集群的完整性和可用性^[15-16]。

2.4 Hadoop 大数据框架

Hadoop 是分布式系统开源框架,核心模块为 HDFS 和 MapReduce。HDFS 为海量的数据提供存储,而 MapReduce 则为海量的数据提供计算。HDFS 有高容错性的特点,并且可以部署在低廉的硬件上,可从单一的服务器扩展到上千台机器,每台服务器都可以提供本地计算和存储。它允许在整个集群使用简单编程模型在计算机的分布式环境下存储并处理大数据,用户可以在规定的软件框架下屏蔽分布式底层细节进行分布式程序的开发^[17]。Hadoop 的 MapReduce 功能实现将单个任务打碎,并将碎片任务(Map)发送到多个节点上,之后再以单个数据集的形式加载(Reduce)到数据仓库里^[18-19]。

3 基础数据支撑平台功能模块建设

3.1 行业气象服务产品数据库设计和搭建

气象大数据云平台“天擎”(CMADaaS)全面继承了 CIMISS 规范、数据种类和接口服务标准,是当前业务流程整体最优、数据服务统一高效的全国气象部门信息化系统的核心^[20-21]。基于分布式数据环境的 MICAPS 系统提供了海量预报数据和数值模式预报产品、集合预报数据供预报员使用。为了对数据源进行统一采集、存储、运算、管理和共享,通过虚拟资源池配置 2 台虚拟机部署 MYSQL 开源数据库用于存储结构化数据,另外 3 台虚拟机安装 Hadoop 大数据框架,利用其核心模块 HDFS 存储非结构化数据,

从而建立以关系数据库为基础、多种数据形式共存的行业气象服务产品数据库。数据库包括数据存储配置表、任务配置表、任务消息表、数据采集任务表、解析任务记录表、专线用户信息表、用户产品清单表、专线用户记录表等数据表,用于存储和管理各子模块所需的公共数据和各类服务产品。

3.2 数据采集解析系统

数据采集解析系统采用 B/S+C/S 混合架构,利用 JAVA 技术开发数据采集、解析等功能模块。针对气象数据和分布在各个部门的业务数据从识别采集到应用方面,采用自动化作业的解决方案,实现气象业务数据和部门共享数据自动识别、自动分类、自动驱动计算模型,进行数据的收集、归档、清洗、质控、解包、打包、格式转换等数据处理的操作,从而不断地提高广西行业气象数据服务能力和质量。所有的自动服务进程均可进行半自动人工操作干预,数据处理模式可切换。数据采集解析服务流程图如图 2 所示,通过配置定时器触发采集需要的气象数据,监听数据采集的情况并利用 kafka 进行消息队列通讯,Zookeeper 分布式协调服务将任务分配给数据解析服务进行解析作业,实现了 http、FTP、文件、数据库、MDFS 等高效采集解析。数据采集解析系统具有任务管理、消息管理、规范管理、系统设置、解析记录等功能模块。

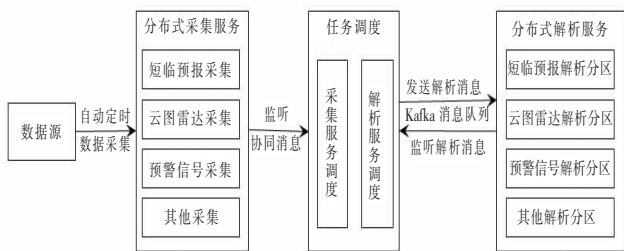


图 2 数据采集解析服务流程图

3.3 大数据推送管理系统

大数据推送管理系统采用 JAVA 等技术开发,通过数据统一服务接口对各个业务系统或专线用户提供数据和产品服务,利用 Redis 缓存技术和分布式推送节点进行产品分发,有利于提高数据推送时效。大数据推送管理系统具有数据推送(推送任务、推送记录查询)、系统管理(专线用户管理、专线用户数据清单)等功能,可实现采集目标(数据库、文件、MDFS)、推送频次、推送时间、推送目的地、推送任务启动/停止的灵活配置。

3.4 数据监控平台

数据监控平台采用 JAVA、WEB 等技术开发,是对平台及体系内各服务、数据和应用的运行情况进行实时监控,具有监控大屏、API 监控、应用程序监控、服务器监控、数据推送监控、数据采集解析监控、全流程监控、异常信息查询、自动告警等功能。

监控大屏通栏各个服务器和系统运行状态、查看实时可视化数据报告。API 接口监控是对 API 接口进行监控,是否有报错等。应用程序监控是监控各业务系统运行情况。服务器硬件监控是监控服务器运行情况,如是否宕机、CPU、内存、网络和磁盘等实时使用情况进行监控。数据推送监控是监控数据推送服务情况。数据采集解析监控是监控数据采集是否采集解析失败。全流程监控是从数据采集、解析、推送进行全流程监控。自动告警功能是通过监控指标(告警规则)、告警渠道(如钉钉群、邮箱、短信等)及联系人的自定义配置后设定自动启动,如监控运行状态出现异常达到告警规则时,立即通过告警渠道将告警信息发送给相应的联系人。

3.5 业务系统综合管理平台

业务系统综合管理平台采用 JAVA、WEB 等技术开发,是针对各类业务服务系统进行统一管理,具有行业管理、系统管理、用户管理、日志管理等功能。行业管理是按照服务行业的不同将业务系统进行分类,并定义行业编码和关注气象要素。系统管理分为平台管理、模块管理、接口管理和角色管理,平台管理主要包括平台编码、平台 LOGO、平台密钥等管理内容,模块管理主要包括模块编码、模块路径、跳转方式、所属平台等内容,接口管理主要包括接口名称、接口地址、请求方式等内容,角色管理根据不同平台不同用户分配不同的角色权限。用户管理主要是针对不同平台的用户进行创建或删除。日志管理主要针对不同平台的操作日志、系统日志、接口日志等进行分类管理。

4 应用情况

2020 年 10 月开始,基础数据支撑平台通过对原有专线用户分散的采集传输程序集约整合,以及对各业务服务系统的数据支撑,实现了气象和行业数据在气象部门内部和外部的共享应用,集成气象服务应用算法,打造了从数据采集加工传输的全流程可视化监控,现已为观天知水气象综合信息系统、广西行业气象服务网站、大藤峡 APP、广西行业气象

影响预报服务系统、广西行业气象服务集约化系统、覆冰专项气象服务平台等业务服务系统提供数据支撑,同时还在电力、水利、自然资源、民航等部门服务中进行业务应用,每日稳定高效地执行采集解析推送任务 45302 项,日数据吞吐量约为 52778MB,极大提高了气象信息的及时性和各行业服务的应用能力。

5 结语

叙述了基础数据支撑平台采用的 Hadoop、Kafka 等关键技术,并对平台构建流程和功能模块进行了细致的介绍。平台实现了在大数据框架下的数据自动采集、解析、推送、监控、业务服务系统管理等功能。

基础数据支撑平台运行正常,已稳定地为各业务服务系统和行业用户提供数据服务,通过全流程可视化监控,提高服务运维能力。

随着气象大数据云平台“天擎”的不断升级完善,以及各行业部门对气象服务要求的不断提高,平台内容及功能仍需进一步改进和完善,以适应社会与经济发展对专业气象服务的需求,进而提供更具针对性的全方位、专业性和精细化气象服务产品,提升行业气象服务保障能力。

参考文献:

- [1] 戚云枫,罗建英,何珊珊.广西决策气象服务系统研发设计[J].气象研究与应用,2020,41(2):55-58.
- [2] 李有华,卢小凤,陈剑飞,等.基于智能网格产品的广西行业气象服务集约化系统的设计与实现[J].气象研究与应用,2019,40(4):59-62.
- [3] 黄丹,植耀玲,伍华丽,等.广西雷电监测预警服务系统的设计与应用[J].气象研究与应用,2021,42(2):73-76.
- [4] 施海瑞,徐爽,刘立辉,等.道路交通气象服务系统的设计与应用[J].地理空间信息,2022,20(3):380-384.
- [5] 黎颖智,李勇,李有华,等.基于 Echarts 的库区水文及气象数据监测系统的设计与应用[J].气象研究与应用,2021,42(2):68-72.
- [6] 杨乐,朱国栋,陈福康.基于网络安全域的民航气象信息服务系统设计[J].民航管理,2019,(6):72-74.
- [7] Hiranman B R, Chapte V M, Abhijeet C K. A Study of Apache Kafka in Big Data Stream Processing [C]//2018 International Conference on Information, Communication, Engineering and Technology (ICICET). 2018:1-20.
- [8] Dobbelaere P, Esmaili K S. Kafka Versus RabbitMQ: A comparative Study of Two Industry Reference Publish/Subscribe Implementations; Industry Paper [C]//Proceedings of the 11th ACM International Conference on Distributed and Event-Based Systems (DEBS'17). Barcelona, Spain: ACM Press, 2017:227-238.
- [9] Delamer I M, Martinez L J L, Perez O. An Evolutionary Algorithm for Optimization of XML Publish / Subscribe Middleware in Electronics Production [C] // Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. 2006:681-688.
- [10] 谢文康,樊卫北,张玉杰,等. ENLHS: 一种基于抽样的 Kafka 自适应调优方法[J]. 计算机科学, 2020, 47(8):119-126.
- [11] 刘邦,余华平. Kafka 分布式消息队列的高性能研究[J]. 电脑知识与技术, 2019, 15(32):4-6.
- [12] 苟丽美,张锋叶,林国华. 基于 Zookeeper 的 GIS 集群实现[J]. 计算机工程与设计, 2017, 38(9):2573-2579.
- [13] 李锐,刘刚,尹书伟,等. 基于 ZooKeeper 和 ACE 的分布式计算框架研究与设计[J]. 电力信息与通信技术, 2015, 13(6):31-34.
- [14] Senn E, Douhib S, Blouin D, et al. Power and Energy Estimations in Model Based Design [M]. Netherlands: Springer, 2009:3-26.
- [15] Wu Z H. Petrinetwork Introduction [M]. Beijing: Mechanical Industry Press, 2006:47-64.
- [16] 姚经纬,杨福军. Redis 分布式缓存技术在 Hadoop 平台上的应用[J]. 计算机技术与发展, 2017, 27(6):146-150, 155.
- [17] 王军,刘文化,于伟东,等. 一种基于 Hadoop 的纺织海量生产数据存储设计[J]. 微型电脑应用, 2013, (6):53-57.
- [18] 王海荣,刘珂. 基于 Hadoop 的海量数据存储系统设计[J]. 科技通报, 2014, 30(9):127-130.
- [19] 任晓炜,黄志,詹利群. Hadoop 在气象数据存储中的应用[J]. 气象研究与应用, 2019, 40(1):65-68.
- [20] 国家气象信息中心. 气象大数据云平台“天擎”1.0 应用开发手册(V1.0) [Z]. 2020:3-4, 64-67.
- [21] 广东省气象探测数据中心. 广东省气象大数据云平台“天擎”应用融入技术指南 [Z]. 广州:广东省气象局, 2021:13-16, 128-131.

Implementation and application of basic data support platform based on industry meteorological services

Chen Jianfei, Shi Caixia*, Wang Zhenguo

(Guangxi Meteorological Disaster Prevention Technology Center, Nanning 530001, China)

Abstract: Based on the industry meteorological services, this paper established the basic data support platform by using Java, WEB and other technologies, and introduced its construction process and functional modules in detail. The system adopts B/S +C/S hybrid architecture, and each layer is independent to improve the security, stability and maintainability. Through the basic data support platform, the data support for each business service system and special line users can be quickly completed, so as to provide reference for improving industry meteorological services.

Key words: industry meteorological services; big data; Tianqing; visualization