

# 基于分布式框架的气象预报服务系统

秦运龙<sup>1</sup> 张冰松<sup>1</sup> 祝 赢<sup>2</sup> 王迎迎<sup>1</sup>

(1.湖北省气象信息与技术保障中心,湖北 武汉 430074;

2.武汉中心气象台,湖北 武汉 430074)

**摘 要:** 随着气象现代化的不断发展,气象服务产品不断丰富,气象观测数据呈现爆炸式增长,各种气象服务系统集成化程度较低,系统并发访问、处理及检索显示效率较低。因此,建立集约、高效的预报服务系统成为业务发展的迫切需求。文中提出了气象预报服务系统,该系统采用 Nginx+Apache 应用方案,Storm 分布式数据处理技术,Cassandra 和 HBase 分布式数据存储技术,RabbitMQ 消息队列统一数据分发技术,分别针对网页负载均衡、动静分离响应,大数据分布式并发预处理,数据分类分布式存储,跨平台、高效、可靠的数据分发等服务进行优化设计,实现了针对长江流域、省、市、县用户的高并发、高处理能力、高可靠性的预报服务系统。该系统的建成极大地提高了气象预报、服务的工作效率,为今后的气象业务发展提供了较好的基础平台。

**关键词:** 负载均衡; 分布式处理; 分布式存储; 消息队列

中图分类号: TP39

文献标识码: A

文章编号: 1673-629X(2019)05-0178-04

doi: 10.3969/j.issn.1673-629X.2019.05.037

## Weather Forecast System Based on Framework of Distributed Services

QIN Yun-long<sup>1</sup> ZHANG Bing-song<sup>1</sup> ZHU Ying<sup>2</sup> WANG Ying-ying<sup>1</sup>

(1.Hubei Meteorological Information and Technical Support Center,Wuhan 430074,China;

2.Wuhan Central Meteorological Observatory,Wuhan 430074,China)

**Abstract:** With the continuous development of the meteorological modernization,the meteorological service products are constantly enriched,the meteorological observation data are increasing with the explosive growth,the intensification degree of various meteorological service system is low,and the efficiency of concurrent access,processing and retrieval display of the system is low. Therefore,the establishment of intensive and efficient forecasting service system has become an urgent need for business development. We put forward a weather forecast service system which adopts Nginx+Apache application scheme,Storm distributed data processing technology,Cassandra and HBase distributed data storage technology and RabbitMQ message queue unified data distribution technology to carry out optimization design respectively for web load balancing,separation of dynamic and static response,big data distributed concurrent preprocessing,classification of data distributed storage,cross-platform efficient and reliable data distribution and other services,realizing the forecast service system with high concurrency,high processing capacity and reliability for Yangtze River Basin,province,city and county users. The construction of the system has greatly improved the efficiency of weather forecasting and service,and provided a foundation for future meteorological business development.

**Key words:** load balancing; distributed processing; distributed storage; message queuing

## 0 引言

“十二五”以来,现代气象业务的快速发展,预报预测精细化水平不断提高,气象观测数据呈现爆炸式增长。湖北作为长江流域气象中心,全省共有预报服务产品30余种,观测站点2600余个,长江流域观测站20000余个,观测粒度由1小时精确到5分钟。针

对观测保障、预报预测、气象服务的业务系统达30余个。气象业务和系统的快速发展同时也使得预报业务和服务需求不协调的问题日益凸显。如何提高业务发展的协调性,系统建设的集约性、高效性、可靠性成为当前气象业务发展迫切需要解决的问题<sup>[1-3]</sup>。

气象预报服务系统采用 Nginx+Apache 应用方案,

收稿日期: 2018-07-12

修回日期: 2018-11-21

网络出版时间: 2019-03-06

基金项目: 中国气象局公益性行业(气象)科研专项经费项目(GYHY201206031); 湖北省气象局科技发展基金重点项目(2017Z05)

作者简介: 秦运龙(1984-),男,硕士,高级工程师,研究方向为气象信息技术。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190306.0952.062.html>

Storm 分布式数据处理技术<sup>[4-5]</sup>, Cassandra 和 HBase 分布式数据存储技术<sup>[6-7]</sup>, RabbitMQ 消息队列统一数据分发技术, 实现针对流域、省、市、县用户的高并发、高处理、高读写、高可靠的预报服务系统, 为气象业务工作提供强大的系统性能支撑、完善的业务逻辑流程、完整的气象业务功能、高效的业务处理服务平台<sup>[8-9]</sup>。

## 1 系统组成

气象预报服务系统标准体系架构采用分层设计思想, 综合考虑数据流程、功能划分, 将系统架构划分为数据展示层、数据分析处理层、数据分发层、数据存储层, 如图1所示。

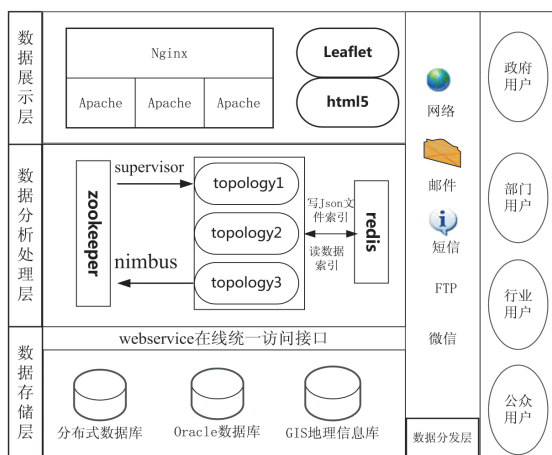


图1 气象预报服务系统总体框架

(1) 数据展示层。数据展示层 WEBGIS 采用 Leaflet 控件通过叠加地理信息展示气象数据。采用 Nginx+Apache 应用方案实现 WEB 访问的负载均衡, 达到可承载 200 人进行极限操作访问的高并发承载能力。

(2) 数据分析处理层。数据分析处理层采用 Storm 分布式数据技术, 实现图片、数值预报产品、气象标准格式数据的快速并发处理。

(3) 数据分发层。数据分发层实现对产品数据的分发管理, 将不同产品采用其相对应的各种分发传输手段快速发送到对应的服务用户, 包括政府用户、部门用户、行业用户、公众用户。

(4) 数据存储层。数据存储层主要采用 HBase 数据库、Cassandra 数据库、Oracle 数据库、地理信息数据存储库, 根据不同数据特性采用不同的数据库进行存储, 实现高效的数据存储读写<sup>[10]</sup>。

## 2 系统设计与实现

### 2.1 数据展示层

数据展示层采用 B/S 架构的 WEB+GIS 页面展示, 通过 Nginx 的轮询方式对 3 台 WEB 服务器做负载

均衡, 增加系统高并发性和可靠性。Nginx 的静态处理能力较强, 但是动态处理能力不足, 因此, 在该系统中采用动静分离技术, 即 Nginx+Apache 应用方案, 静态页面由 Nginx 处理, 动态页面由 Apache 处理。在 Nginx 实现三台服务器负载均衡后, 部署配置相应的 Apache 页面和 Nginx 页面。用户向 Nginx 请求静态数据, 则 Nginx 直接应答, 当向 Nginx 请求动态数据时, Nginx 将请求转发到 Apache 页面, 通过这种代理方式, Apache 提供动态数据请求的 WEB 服务。该方案极大地提升了网站高并发性、稳定性和用户体验<sup>[11]</sup>。

### 2.2 数据分析处理层

气象预报服务系统基于 Storm 分布式框架, 实现实时快速的分布式数据处理, 保证数据的处理和检索效率。在 Storm 分布式框架中, nimbus 部署在主节点上负责在集群中分发代码, 对节点分配任务, 并监视主机故障, nimbus 通过在 zookeeper 上写状态信息来分配任务。supervisor 部署在各工作节点上, 负责监听工作节点上已经分配的作业, 启动和停止 nimbus 已分配的工作进程, supervisor 通过从 zookeeper 上读取这些状态信息领取任务, 如图2所示。

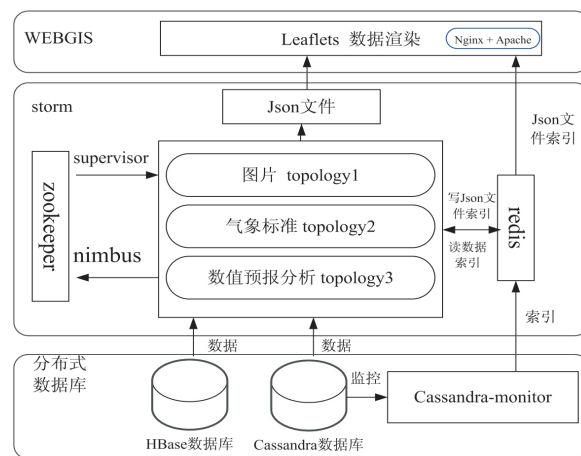


图2 基于 Storm 的分布式数据处理结构

Cassandra 和 HBase 分布式数据库为 Storm 框架提供数据源。Cassandra 主要包括各种数值预报、高空、地面、雷达等非关系型数据和产品, 通过 Cassandra-monitor 接口, redis 实时获取数据索引, 并读取数据, 交由多个并发的 topology 任务对象进行处理; HBase 主要包括湖北省、华中区域、长江流域的分钟、小时、日值气象观测数据, 不同类型数据采用不同的分布式数据库存储, 极大地提高了系统数据的并发处理能力<sup>[12-13]</sup>。前端 WEBGIS 通过 redis 并发检索预处理后的 json 格式文件, 快速显示数据产品, 保证了用户体验。Storm 部分配置代码如下:

```
.....
storm.zookeeper.servers:
- "10.104.xx.51"
```

```

nimbus.seeds ["storm51", "storm52", "storm53"] # nimbus
集群
storm.zookeeper.session.timeout: 200000 # zookeeper 会话的
过期时间 默认值 10 倍
nimbus.task.timeout.secs: 300 #30* 10
nimbus.supervisor.timeout.secs: 600 #60* 10
ui.port: 9999# UI 端口
supervisor.slots.ports:
- 6700
- 6701
- 6702
- 6732
supervisor.worker.timeout.secs: 300 #30* 10 默认值 10 倍
worker.heap.memory.mb: 51200 # worker 内存分配 50G
.....

```

### 2.3 数据分发层

气象预报服务系统数据统一分发模块的设计提高了产品分发的时效性、多样性和可靠性。

数据统一分发基于生产者消费者模型实现,采用 RabbitMQ 消息队列机制,对生产者与消费者实现异步与解耦。此外,RabbitMQ 消息队列具有跨平台、健壮、稳定、易用且有消息确认机制和持久化机制、可靠性高等优点,适用于数据分发任务管理的业务场景,具体功能设计如图 3 所示。

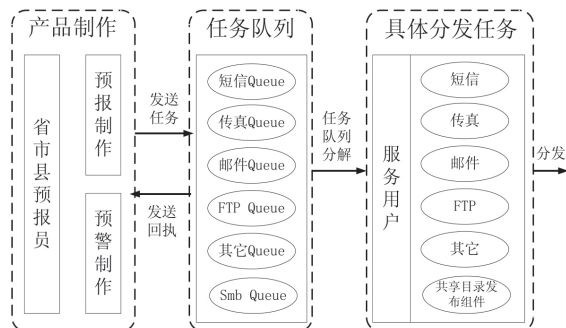


图 3 数据分发管理功能设计

分发任务管理对省市县各级业务系统提交过来的发布任务进行管理,任务按照发布驱动的不同(短信、传真、邮件、FTP、大屏、大喇叭、电话、微信、微博)分别

进入对应消息队列,按照先进先出原则向服务用户分发数据和产品。

### 2.4 数据存储层

预报服务系统管理的数据包括观测资料(自动站、雷达、卫星、闪电、大气成分、酸雨、负离子、雷达产品数据等)、数值预报、产品、行业数据(水文、电力、环境等部门的行业数据)以及系统公共管理信息。由分布式数据库和 Oracle 关系数据库根据数据类型进行分类存储,并且通过 webservice 在线统一访问接口提供数据访问服务,保证数据库访问的稳定性,其结构如图 4 所示。

(1) Oracle 关系数据库:为保证系统业务的延续性和数据标准的统一规划,Oracle 数据库存储建站以来的所有历史数据。

(2) HBase 分布式数据库<sup>[14]</sup>:为保证系统加载实况数据的响应速度,将长江流域、华中区域自动站(国家站、区域站)实时 5 分钟和小时数据存储于 HBase 分布式数据库,保存 6 个月。

(3) Cassandra 分布式数据库<sup>[15]</sup>:为提高数值预报模式的加载效率,将采用 Cassandra 数据库作为数值预报、雷达、高空和地面填图产品的直接数据源,数据保存 15 天。

## 3 系统分析

气象预报服务系统以 Java 为主要开发语言,采用 B/S 架构,模块化设计各功能。通过 Storm 分布式技术和最新的 html5 语言,实现良好的用户交互体验。

### 3.1 系统功能分析

相对于其他系统,文中建设的系统业务覆盖范围更全面,服务业务人员更广泛,处理响应性能更安全可靠。其分布式框架技术极大地提高了用户高时效、高并发时的良好操作体验。基于 RabbitMQ 消息队列的数据统一分发技术为业务人员提供了更稳定、更多样的多目标多渠道分发服务。与其他系统的对比如表 1 所示<sup>[16-17]</sup>。

表 1 系统对比

对比项	其他系统	文中系统
数据展示	单机 html、PHP	3 台负载均衡、Nginx+Apache 动静分离
数据存储	分布式或关系数据库存储	分布式+关系数据库
数据处理	常规单机处理	基于 Storm 的分布式快速处理,html5 交互技术
业务范围	单一预报	监测、预报、预测、服务(专业、专项、为农、公众、决策、流域)
业务制作	单独制作	省市县业务联动交互,信息反馈
数据分发	单用户、单一方式发送	基于 RabbitMQ 消息队列多用户、多渠道并发推送
服务对象	本级业务人员	流域、省、市、县四级业务人员

### 3.2 系统性能分析

系统性能测试主要表现在大数据量情况下访问系统时随着数据量增加响应时间的变化,对于类似气象这种拥有大数据吞吐量的行业,其响应时间对于用户体验至关重要。文中测试数据为长江流域 20 000 个观测站小时降水数据,数据库分别为分布式数据库和传统 Oracle 数据库,存储数据量为历史所有气象数据。测试系统为传统单网页系统和文中分布式框架系统。通过分别加载 00 点-12 点 12 个时次长江流域降水数据统计,对比响应时间,如图 4 所示。

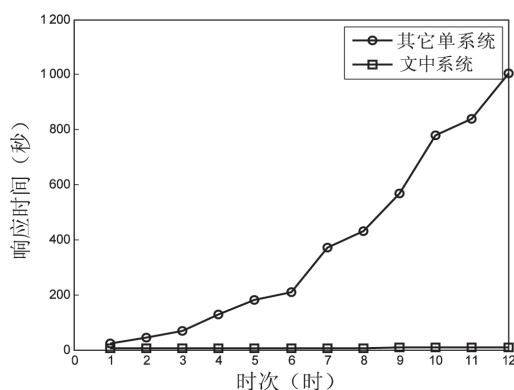


图4 系统响应时间对比分析

可以看出,随着统计数量的增加,文中系统响应时间均在 2 s 以内,但传统系统响应时间急剧上升,无法满足现有气象大数据量应用的需求。

## 4 结束语

提出了一种基于分布式处理框架,采用 Cassandra、HBase 分布式存储技术的气象预报服务系统,极大地提高了用户高并发时的良好操作体验,实现了流域、省、市、县用户多功能、高效率、高并发、高可靠的集约化系统,解决了各部门间重复建设问题,提高了部门之间的信息交互和业务协调能力,为今后的信息化、智能化建设提供了较好的基础和保障。

### 参考文献:

- [1] 孔 钦,叶长青,孙 赞.大数据下数据预处理方法研究[J].计算机技术与发展,2018,28(5):1-4.
- [2] MITRA P,MURTHY C A,PAL S K.Density-based multi-scale data condensation[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,24(6):734-747.
- [3] BACARDIT J,WIDERA P,CHAMORRO A E M,et al.Contact map prediction using a large-scale ensemble of rule sets and the fusion of multiple predicted structural features[J].Bioinformatics,2012,28(19):2441-2448.
- [4] 蒋晨晨,季一木,孙雁飞,等.基于 Storm 的面向大数据实时流查询系统设计研究[J].南京邮电大学学报:自然科学版,2016,36(3):100-105.
- [5] LI G J,CHENG X Q.Research status and scientific thinking of big data[J].Bulletin of the Chinese Academy of Sciences,2012,27(6):647-657.
- [6] 李永生,曾 沁,徐美红,等.基于 Hadoop 的数值预报产品服务台设计与实现[J].应用气象学报,2015,26(1):122-128.
- [7] 程 苗,陈华平.基于 Hadoop 的 Web 日志挖掘[J].计算机工程,2011,37(11):37-39.
- [8] 李 洋,吕家恪.基于 Hadoop 与 Storm 的日志实时处理系统研究[J].西南师范大学学报:自然科学版,2017,42(4):119-126.
- [9] 陈庆奎,周利珍.基于 HBase 的大规模无线传感网络数据存储系统[J].计算机应用,2012,32(7):1920-1923.
- [10] 贺雅楠,高 嵩,薛 峰,等.基于 MICAPS4 的智能网格预报平台设计与实现[J].应用气象学报,2018,29(1):13-24.
- [11] 瞿龙俊.基于 HBase 的交通流数据实时存储与查询优化方案的设计与实现[D].镇江:江苏大学,2017.
- [12] 王建荣,华连生,唐怀颐,等.数值预报产品分布式处理与存储系统设计[J].计算机技术与发展,2018,28(2):167-172.
- [13] 朱 涛,郭进伟,周 欢,等.分布式数据库中一致性与可用性的关系[J].软件学报,2018,29(1):134-149.
- [14] YANG Wenjie,LIU Xingang,ZHANG Lan,et al.Big data real-time processing based on storm[C]//IEEE international conference on trust,security and privacy in computing and communications. Melbourne, VIC, Australia: IEEE,2013:1784-1787.
- [15] 顾文静,常 飏.集约化数值预报业务自动化管理系统可视化平台的设计与实现[J].气象科技,2014,42(2):266-272.
- [16] 朱静静,黄海于.分布式仿真平台用户端工况管理的设计与实现[J].计算机技术与发展,2018,28(2):135-139.
- [17] 齐方方.海量数据存储和准实时查询系统设计与实现[D].南充:西南石油大学,2015.