

多渠道气象服务数据在线平台的设计

■ 胡娟 李游 张立杰 张凯 郑慧

采用移动互联网、云计算、大数据技术，通过对气象服务大数据挖掘，对“深圳天气”品牌多渠道气象海量数据进行挖掘及服务产品实时监测与分析，实现数据驱动再生，提出了“深圳天气”服务品牌在线监控与服务数据抓取的技术指标，并初步设计实现了多渠道气象服务在线管理平台。

DOI: 10.3969/j.issn.2095-1973.2019.03.037

传统的气象服务方式是主动为用户提供气象信息服务，其缺点是服务方式单一，缺乏多方的参与。随着移动互联网、云计算、大数据技术的出现，深圳市气象局以“深圳天气”为品牌陆续推出了微网页、微博、智能手机客户端、微信公众服务平台、深圳台风网、数据开放与共享平台等一系列渠道服务，充分满足深圳市民对气象服务的多样性、个性化的需求，为实施科技气象战略，推进精细化、科学化管理，提高气象公众服务的品质和水平起到了一定的作用。开展多渠道气象服务数据在线管理平台建设，利用信息技术整合以“深圳天气”为品牌的公众服务多渠道数据，实现各“孤立”系统的服务数据共享，对各渠道数据进行统计分析，可以为提升“智慧气象”服务水平提供决策参考。

1 多渠道气象服务管理需求分析

1.1 多渠道气象服务数据在线管理应用场景分析

多渠道气象服务在线管理平台以公有云存储为基础，以云计算为处理核心，建立统一规范化数据格式，形成以海量数据业务支撑的大数据平台，逐步形成气象信息服务“云+端”的运行模式，并在实时处理领域实现秒级突破。对微博、微信公众号（订阅号）、智能手机客户端、WEB网站数据进行统计，可以实现对“深圳天气”各个渠道业务数据实时查看与统计，方便决策者和气象服务人员快速做出响应，以适应移动互联网服务对时效性的要求，而传统监控基本都是对年、月、周、日的频次进行统计，已经不能满足个性化的服务需要，现今需要实现24 h在线实时监控以及管理当前实时变化的统计数据，最终实现7×24的用户实时行为监测和秒级分析。建立各渠道

接入站点的运营监控流量指标如UV（访客数，在同一天相同的客户端只被统计一次）、PV（页面浏览量或点击量，刷新一次统计一次）、IP（独立IP，同一天内相同IP统计一次）、新旧访客数等，同时，为提供更好的服务质量，建立服务质量指标体系，如用户访问停留时间、访问次数、访问深度、跳出频次等，对此进行立体式汇总与分析，如平均停留时间、平均加载时间、跳出率等以方便进行全局分析，提升产品的质量与体验度，增强用户的使用粘度。加入一定的用户的行为分析、网站访客背景分析、鼠标点击行为等高智能的分析功能，为智慧气象服务提供决策数据支撑。

1.2 数据处理及信息流机制相关分析

通过对气象大数据平台的引入，标准化数据处理格式，统一数据接入，打通气象业务系统与服务渠道之间的数据交换及数据格式的转换，建立一体式气象业务服务总线，对业务数据进行多渠道、多维度、精细化分析，挖掘出有价值的业务数据信息，以应对移动互联网下的“智慧气象”的需求。多渠道气象服务数据在线管理平台支持多种风格的数据展现方式，使用当前流行的线性、柱形、饼状分析图，对不同时段、不同服务渠道及服务访问量分类展示，引入创新风格的热力图，更加视觉化的展示多渠道气象服务页面不同位置访问的密度，增强气象产品服务特色体验，并能实时反映出公众感兴趣的气象产品，采用大数据屏幕方式展现气象重要业务及实时数据，既符合现代流行的大数据展现方式，同时也方便气象服务产品的服务监督。

收稿日期：2018年12月3日；修回日期：2019年4月10日
通信作者：李游（1983—），Email: 174069267@qq.com

2 多渠道气象服务数据在线管理平台设计

2.1 平台基本功能

多渠道气象服务数据在线管理平台以深圳气象大数据平台为依托，以气象大数据分析为导向，利用大数据分析工具，通过获取微博、微信公众号（订阅号）、智能手机客户端、WEB网站等渠道获得实时的气象数据及服务数据。平台应具备几个功能点：1) 多渠道、多维度分析展现气象产品服务访问指数，并按移动互联网主流的图表进行展示；2) 根据用户的特点分析出公众感兴趣的气象服务产品，以公众为导向，智能推送其感兴趣的气象服务产品，开展基于位置服务精细化智能气象服务；3) 重要业务数据及实

时数据以大数据服务监督展示，数据展现方式以通俗、易懂、易用为主，具有一定的科技性；4) 智能移动终端可访问平台开放的气象服务后台；5) 结合智能决策库，智能的提供气象服务产品，对用户提出与气象相关的问题做出科学合理回复；6) 对各个气象服务渠道按日、周、月进行用户访问行为分析，总结气象服务的情况，自动生成日、周、月报；7) 制定机器学习规则，分析出历史气候和当前的气候变化，提供科学性分析与总结。

2.2 平台体系结构

多渠道气象服务数据在线管理平台总体设计框架（图1）在层次上可划分为基础层、应用支撑层、数据层、应用层和用户层。

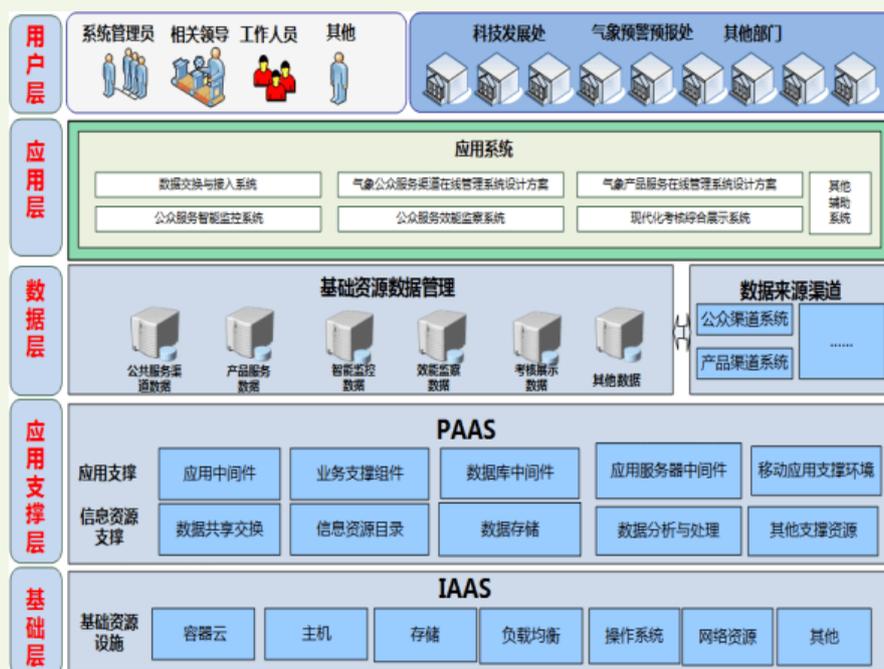


图1 多渠道气象服务数据在线管理平台结构图

基础层：基础层即平台所需的基础设施，主要包含主机、存储、操作系统、网络资源、容器云等基础资源。

应用支撑层：应用支撑层即平台运行所需的支撑组件，主要包含应用中间件、数据库中间件、数据共享交换组件、数据存储组件等。

数据层：数据层即平台运行所需的基础数据资源，主要包含公共服务渠道数据、产品服务数据、智能监控数据、效能监察数据、考核展示数据等系统所需的数据。

应用层：应用层即平台建设实际展示和应用的系统，主要包含数据交换接入系统、气象公众服务渠道

在线管理系统、气象产品服务在线管理系统、公众服务智能监控系统和公众服务效能检查系统等。

用户层：用户层即系统具体涉及的用户，主要包含系统管理人员、相关领导、工作人员等。

2.3 系统整体架构及技术选型

基于微服务的设计思路，将业务系统按功能拆分成多个独立的服务，如用户统一授权及管理服务、气象核心产品服务、后台管理服务、海量数据分析与检索服务，分布式调度服务，分布式文件存储服务等等。这些独立的服务运行在Docker容器中，各个服务之间通过消息中间件（例如KafKa）或RPC通信，服务治理使用Zookeeper。

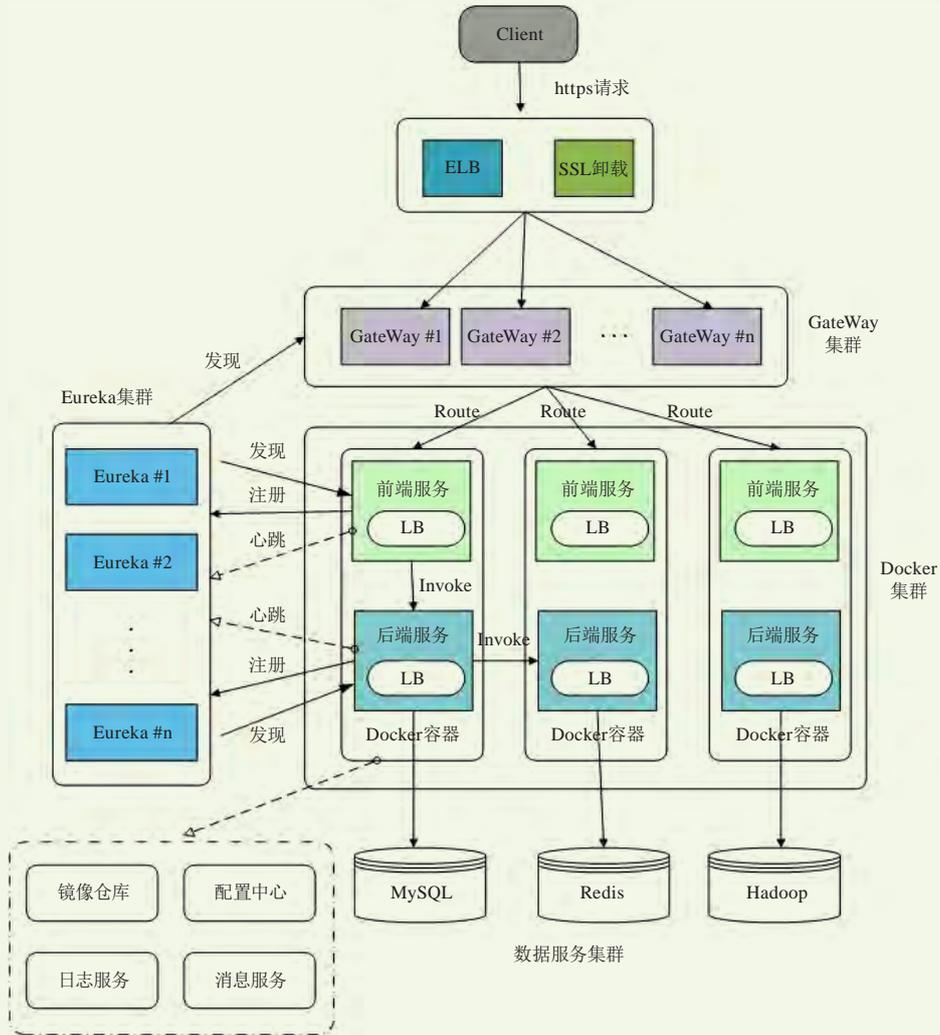


图2 多渠道气象服务数据在线管理平台微服务架构

使用这种架构（图2）主要基于以下几点考虑：

- 1) 系统之间的耦合度（包括服务层与表现层）大大降低，各个服务可以独立开发、独立部署、独立测试，服务与服务之间边界明确；
- 2) 系统利于横向和纵向扩展；
- 3) 系统中流量分散，可承载更高的负载；
- 4) 服务的复用性更高，各个系统可以共享服务。

2.4 实现海量数据处理及存储

首次应用气象大数据技术进行设计以适应未来气象业务的高速发展，使用分布式数据库HBase对海量数据存储，可以提高数据的并发访问量，传统的关系型数据库提供读写分离的方式，但是带来数据延时；对于越来越多的海量数据，传统数据库采用的是分库分表，实现起来比较复杂，后期需要不断进行数据迁移维护；高可用和伸缩性方面，传统数据库采用的是主备、主从、多主的解决方案，增加节点和宕机需要

进行数据迁移。对于上述问题，分布式数据库HBase有一套完善的解决方案，适用于高并发海量数据存取的需求。使用分布式计算对业务数据进行分析计算，使用ElasticSearch搜索引擎对海量数据进行检索，并以图表形式展现出来。

2.5 实现多种并行展示方式

通过互联网等信息传播途径，当用户通过浏览器浏览网站或网页时，与后台的WEB服务器和气象信息数据库服务器不断进行通信，根据不同的用户的不同需求，系统提供以文本、地理信息图像、三维动画、智能手机客户端等多种方式将所需信息展现出来，满足不同用户的个性化需求。

- 提供全新的气象服务模式，通过人机互动，以用户需求、用户习惯为导向为公众提供个性化气象服务；
- 为用户提供实时的气象信息服务，提供多维度

多渠道数据分析:

- 提供更精确的气象数据, 让数据更及时, 更精确, 适应气象日益增长的气象业务管理需求;

- 提供多渠道的服务方式, 用户可通过智能手机客户端和互联网络等多种方式获得精细化的气象信息服务;

- 通过机器学习, 记录用户操作行为习惯而提供其关注的服务内容及数据, 逐步形成气象的专家系统, 提供更加智能的气象服务;

- 健全电子考评机制, 提供科学化的考评方案;

- 在应对重大突发事件时, 更好、更及时、更精准地为公众提供及时、准确的信息服务, 同时为领导提供科学的公众气象决策支持服务。

2.6 平台涉及的关键技术

多渠道气象数据在线管理平台的建设涉及到海量数据与产品, 管理的用户数据是千万量级, 不同的渠道技术架构完全迥异, 因此建立多渠道气象数据在线管理平台需要突破以下关键技术: 1) 基于网络技术开发专用网络爬虫, 在不影响渠道服务性能的情况下采集用户信息及服务信息, 无需修改渠道源代码, 为用户行为分析奠定数据基础; 2) 利用AI技术开发客户行为分析模型, 为智能推送服务提供基础支持; 3) 基于并行计算技术, 实现多源数据采集与同步推送, 实现海量数据再分析与展示; 4) 基于可视化技术和消息树技术解决前台展示难题, 实现系统不间断运行与监控, 并提升用户体验。

3 展望

通过多渠道气象数据在线管理平台的建设, 将有助于提高公众获得信息资源的能力, 提供个性化, 多样化的气象信息的服务, 及时满足公众的气象信息的需求, 对于当今“智慧城市”的建设, 乃至“数字政府”的建设起到里程碑的作用。目前, 深圳市已实现

突发事件预警信息的13种发布渠道的一键式发布, 预警信息1 min之内可服务于公众。实现手机预警短信面向公众最小以街道为单元的分区分精准靶向发布, 全网短信发布从原来的8~9 h 提升至2~3 h完成, 覆盖来深漫游用户。实现重大天气电视台整点播报, 平安大厦、京基100等高楼外墙视频播放台风预警信号, 将台风预警信息推送至全体市民的机顶盒开机页面。

目前多渠道气象数据在线管理平台还存在很多不足, 这些不足和改进的思路主要表现在以下几个方面: 1) 缺乏完备的数据流监管与告警, 以雷达图为例, 作为一份公共产品被各渠道引用, 产品一旦在某个环节出现问题, 由于加工与传输的链条较长, 人工快速定位与排除故障周期太长, 很容易降低用户体验, 因此需要在平台对各产品从原始数据获取、产品加工、产品传输、前台调用各环节进行状态监管并及时告警, 迅速给出故障定位; 2) 基于大数据的客户画像还停留在原始阶段, 各渠道的用户行为数据没有集中存储, 客户画像的特征向量没有经过全面的客观评估与检验, 下一步需要进一步统合用户行为数据, 建立统一的客户行为机器学习模型, 在实践中加以改进; 3) 平台还需要在评选用户最喜欢的产品与服务上下功夫, 按照客户类型、时间段、地点等分类动态评选客户最喜欢的产品排名, 为管理者优化各渠道的产品和服务提供决策依据。

深入阅读

- 贺佳佳, 吴蔚, 孙石阳, 2010. 深圳大运会公共气象服务建设刍议. 广东气象, 32(3): 49-51.
- 张承福, 1994. 神经网络在天气预报中的应用研究. 气象, 20(6): 43-47.
- 章国才, 2006. 防御和减轻气象灾害. 气象, 3(3): 3-5.

(作者单位: 深圳市气象局)