

王垒. 面向服务的气象大数据 [J]. 陕西气象, 2016 (4): 36-39.

文章编号: 1006-4354 (2016) 04-0036-04

# 面向服务的气象大数据

王 垒

(陕西省气象信息中心, 西安 710014)

**摘 要:** 依据国家大数据发展战略和大数据特性, 对气象大数据特征、面临问题和技术创新进行阐述, 深入分析气象大数据服务方式和跨领域服务的发展方向。提出以标准先行为前提, 集约整合为手段, 通过数据层、逻辑层和界面层的整合来搭建面向服务的气象大数据环境, 借助大数据发展机遇加快推动气象大数据发展步伐。

**关键词:** 气象大数据; 数据服务; 数据整合

**中图分类号:** P413

**文献标识码:** B

近年来, 信息化浪潮席卷全球, 尤其以互联网为核心标准的信息技术和产品正在深刻影响社会发展。数据与物质、能源一样都是最宝贵的生产要素, 是战略性资源。国务院关于印发促进大数据发展行动纲要的通知国发〔2015〕50号)明确提出: “2018年底前建成国家政府数据统一开放平台, 率先在信用、交通、医疗、卫生、就业、社保、地理、文化、教育、科技、资源、农业、环境、安监、金融、质量、统计、气象、海洋、企业登记监管等重要领域实现公共数据资源合理适度向社会开放, 带动社会公众开展大数据增值性、公益性开发和创新应用, 充分释放数据红利, 激发大众创业、万众创新活力。”大数据技术是对海量数据的高效存储和处理。目前气象数据增量达 1 PB/年左右, 数据类型复杂, 呈现出大数据的特性。气象信息化的发展为气象领域大数据奠定的基础。

## 1 气象大数据

### 1.1 大数据

何谓大数据, 目前业界还没有公认的说法。就其定义而言, 大数据是一个较为抽象的概念。广义上讲, 大数据包括大数据技术、大数据工

程、大数据科学和大数据应用等相关领域。对大数据进行广义分类是为了适应信息经济时代发展需要而产生的科学技术发展的趋势<sup>[1]</sup>。

大数据是以 TB 为基本计量单位, 以年增长量至少翻番的互联网动态三维数据, 三维数据包括个人、社会和物质世界数据。

大数据阶段是经历数据、信息、知识迭代过程演化而来, 同时大数据在处理能力和手段上较原始分析处理方法有质的飞跃。由于大数据所属领域不同, 其查询及处理需求的分类不同。例如, 互联网行业按照其业务需求, 可以将大数据处理技术分为在线、近线以及离线。其中, 在线模式数据的处理时间一般限定在 ms 甚至是  $\mu\text{s}$  范围内, 而离线模式数据的处理时间可延长至以 d 为单位, 近线模式的数据处理时间则界于二者之间; 而按照处理需求划分, 大数据的处理需求有可面向于海量数据的分布式处理、非结构化数据处理以及实时数据处理。

目前大数据发展呈现数据从多聚类、少融合, 向聚类与融合并重发展, 发展趋势可归纳为: 大数据即服务, 为某一环节或某一问题提供解决方案; 不求所有, 但求所用—参与跨行业的

**收稿日期:** 2015-02-26

**作者简介:** 王垒 (1980—), 男, 陕西蓝田人, 汉族, 硕士, 工程师, 从事信息化和大数据方面工作。

**基金项目:** 陕西省气象局科技创新基金项目 (2015M-55)

数据交易; 精准甄别, 特征画像—改进业务决策模式, 提升竞争力; 打通信息闭环, 建立与客户的产业互动; 构建大数据专业化机构, 细分行业市场, 形成良性生态; 做好大数据使行业市场更加细分和碎片化的应对; 参与智慧城市建设, 提升行业大数据与政府大数据的社会供给融合。

## 1.2 气象大数据

气象部门的信息化建设在我国起步较早, 发展也很迅速, 经多年努力取得不少成绩。目前气象部门每年新增数据量达到 1 PB, 传统的数据存储和处理方法越来越吃力, 气象大数据局面日益形成。如何运用大数据分析对气象数据进行处理是一个急需解决的问题。

### (1) 数据特征

气象大数据主要分为气象领域大数据和气象互联网大数据。气象领域大数据包括气象观测数据(部门内外)、业务派生数据和中间产品数据、职能部门管理数据(财务、人事、项目……)和各类业务系统的状态数据和日志数据等。气象大数据采集不仅包括传统的气象数据, 还包括其他可能用到的数据<sup>[2]</sup>。气象互联网大数据包括移动端搭载的气象要素传感设备的探测数据、用户随手拍摄上传的天气状态照片、搜索引擎对气象相关敏感词的统计分析数据和其他所有可供气象部门应用的互联网数据。两者具体差别如表 1 所示。

表 1 气象大数据对比

数据类型	气象领域大数据	气象互联网大数据
数据来源	部门内部	互联网、众筹
要素内容	专业、全面	简单
时空密度	专业化, 均匀	不均, 极密或极疏
要素精准度	精准	参差不齐
传感器载体	专业探测设备	移动终端、家用电器、交通工具、非气象监测设备等
获取代价	国家财政	免费
体量	较大, 可预测	不详, 未来巨大

### (2) 面临的问题

气象大数据发展面临的问题: 一是部门内数据分散、信息孤岛严重; 二是与气象有关的数据收

集缺乏全面性和系统性; 三是大数据应用能力严重缺乏, 没有大数据处理、分析、挖掘、可视化等方面的技术储备和人才储备; 四是数据开放共享、安全保障、资产保护等方面顶层设计的缺失。

气象大数据建设面临的问题: 一是数据整合难, 数据质量较差, 数据标准不统一; 二是平台整合难, 平台条件不一、标准不同; 三是技术综合难, 技术特点不一, 技术标准不同; 四是需求适应难, 需求变化多样, 需求标准不一。

### (3) 特点及技术创新

大数据应该具备 5V 特征, 即数据规模大 (Volume)、数据种类多 (Variety)、处理速度快 (Velocity)、数据价值密度低 (Value)、数据真实性 (Veracity)。气象数据基本具备大数据特征, 同时气象数据可以作为一个重要的维度, 参与各类大数据产品的发放和服务。气象数据时空内容丰富, 依靠深度的数据挖掘和机器学习, 将气象数据的价值渗透到其他行业的数据分析中, 可以实现具有气象特色的大数据价值最大化。

大数据也在挑战着技术创新, 尤其是大量与业务逻辑和技术无关的数据处理技术, 主要表现在以下 4 个方面: 信息展现技术, GIS 地图、轨迹图、热力图、辐射图和标签云图等; 分析算法, 社会网络、自然语言处理、时序分析和逻辑回归等; 处理架构, Hadoop 分布式计算平台、SQL-on-Hadoop、Spark 和 STREAM 实时计算等; 数据获取, 网页埋点、网络爬虫、分布式数据存储和传感器技术等。这些技术也是气象大数据发展过程中需要解决和掌握的。

## 2 气象大数据服务方式和方向

### 2.1 服务方式

(1) 部门内部数据服务 部门内部数据服务主要以基础数据和数据产品的方式提供, 这些数据主要来源气象业务和气象领域内相关平台。气象业务包括观测业务, 涵盖地面、高空、辐射、雷达、卫星和环境等实况观测数据; 天气业务, 涵盖主观预报、数值预报等预报相关产品; 气候业务, 涵盖气候预测、气候变化、气候模式等气

候相关产品；卫星业务，涵盖静止卫星、极轨卫星以及空间天气相关产品；服务业务，涵盖高空、海洋、辐射等 14 大类基础数据及数据产品服务。气象业务数据也包括数据归档及数据备份恢复系统、产品实时模式与手工模式展示、规范化和标准化产品生成处理流程、评估与报警机制、信息全流程监控和服务产品发布平台等生产的数据产品和日志数据。

(2) 行业数据服务 行业数据服务主要是面向数据服务接口的软件开发及应用，气象部门非公开接口是其他行业基于气象因素提供公共服务的重要资源，针对非公开接口，仍处于通过气象部门的业务推广，双方设立项目的形式开展。由于气象部分的业务推广和合作的技术瓶颈，很多数据合作仅局限于数据推送层面。由于对行业之间数据服务的限制，因此对气象数据价值的深层挖掘和增值服务开发还很欠缺。

## 2.2 服务方向

在国家重大方针政策的引导下，加大气象数据资源与行业数据资源、生态数据资源和社会经济数据资源的关联分析和融合利用，完善对经济的动态监测、开放共享、吸引众智、敏捷响应社会需求，用气象大数据手段洞察民生气象需求，优化配置气象服务资源，提升公共服务水平，提高民生气象服务的在线化、便捷化、均等化水平，为人民生活提供无微不至的个性化气象服务，为社会生产提供专业化的气象服务。

(1) 防灾减灾领域 通过挖掘气象大数据发现灾害产生的各种气象因素，及时完善各类气象灾害临界指标、借助气象大数据分析提高气象灾害预警预报的精准度和时效性，做到完善灾害防御体系、提高灾害预防能力和提升应急救灾能力。

(2) 旅游领域 定制天气气候景观和专项旅游产品，在经过数据深度挖掘、分析、继续加工，形成全国范围、围绕某一气象要素的动态产品地图，结合旅游数据进行深度分析，提供吃、穿、住、行的综合服务，根据游客热力分布、团队位置定位、景区关注度等旅游数据进行分析，对游客聚集多的景区或路线进行优先发布预警。

(3) 交通领域 通过加强对交通高影响天气

的监测和预警，为高速公路、铁路、航运、航空等部门制作针对性的专业气象交通服务产品，进一步结合交通视频等数据，为公众推荐最佳路线选择，提升公共交通气象服务能力。

(4) 农业领域 开展基于气象大数据的农业天气预报，精准地服务于各地农业生产，通过气象大数据实现产量预报、气候品质评价和营销决策等。

## 3 面向服务的气象大数据环境

面向服务的气象大数据环境首要解决基础标准问题，即气象大数据基础性数据资源标准规范、基础设施资源池标准规范、气象模型和算法通用标准规范和气象数据服务的通用标准规范。同时采用集约整合的思想来构建开源、开发、构件和接口的气象大数据环境。

### 3.1 设计思路

气象大数据环境搭建以气象数据整合展开，包括：数据层整合、逻辑层整合和界面层整合。数据层的整合分为：数据库的整合、文件数据的整合、实时数据的整合、硬件和系统信息的整合。逻辑层的整合分为：接口式整合（被整合系统提供服务接口）、入侵式整合（被整合系统没提供接口）。界面层的整合分为：对于 B/S 和 C/S 展示系统整合。气象大数据环境实现路径如图 1 所示。

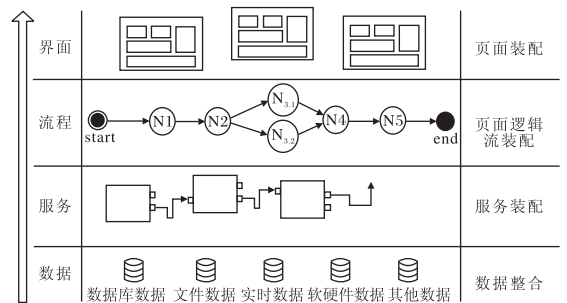


图 1 气象大数据环境实现路径

### 3.2 实施方案

数据层的整合是对数据源的整合，数据底层整合，灵活性高，但整合难度较大。针对数据源的整合其前提是被集成整合的数据源必须开放访问且公开数据结构，即对于关系型数据库必须公开表结构，表间关系，编码的含义等；对于非结

构化数据必须公开数据存储类型和读取方法；对于文本数据必须公开文本存储结构，字段意义，数据格式等；对于实时数据必须公开数据流量大小，数据解密方式，数据传输频率等信息。整合方式是运用多数据源适配技术，在运行监控系统中建立原始数据整合逻辑，该逻辑可适配各种数据源类型，对不同的数据格式进行适配处理。数据层整合逻辑图 2 所示。

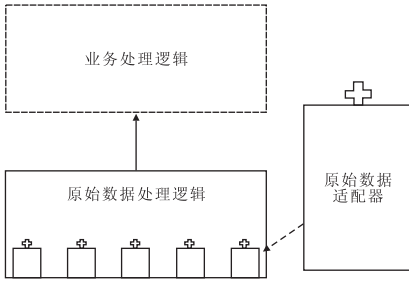


图 2 数据整合逻辑图

逻辑层的整合是数据业务加工的整合，主要是业务逻辑的整合，关键点是协议转换和服务装配。整合前提是整合的业务系统提供用于监控的接口或服务，或可通过二次开发具备提供接口或服务的能力。业务逻辑整合有 3 方面：①业务逻辑层的整合建立在企业服务总线（ESB）和服务装配的基础上；②通过协议转换同各业务系统交互；③ESB 提供的服务，按照业务需求装配编排，可适应未来业务系统的变更、整合和扩展。通过逻辑层整合达到化繁为简作用。

界面层的整合是数据展示的整合。关键点包含界面构件、界面构件库、界面可视化装配环境的建立。整合前提是整合业务系统无相应的服务接口可以使用，且无法为其他业务系统提供数据支持。这样可以安全、稳定的集成现有系统，保留原有监控软件资产。界面层整合包括 4 种方式：①建立界面构件，界面构件可嵌入 C/S 客户端、网页、图表、数据列表等；②建立界面构件库，构件库按照基础构件、容器构件、业务构件进行搭建管理；③监控点的增加就是界面构件

的添加，可为未来的监控业务扩展带来便利；④界面可视化装配环境提供通过拖拽方式进行监控界面的按需定制组装，做到一屏展示多个监控点。界面层整合示意如图 3 所示。

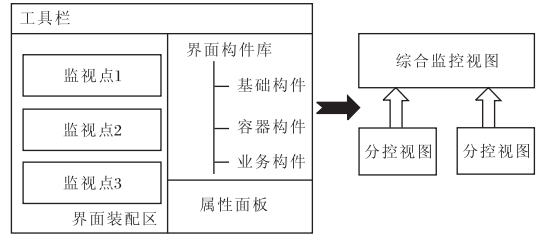


图 3 界面层整合示意图

### 4 结语

实施气象大数据，是适应大数据发展趋势，也是落实国家大数据战略，对促进气象与经济社会融合、强化气象行政事中事后监管、促进气象预报准确率的提高都具有十分重要意义。气象大数据是实现气象信息化的核心关键，因此在气象部门应该推进气象大数据战略，加强气象大数据应用能力、安全保障体系、人才队伍建设，构建健康可持续发展的气象大数据生态系统，推动气象与经济社会的融合，提升气象在国民经济中地位和作用。由于气象部门以大数据技术为核心的信息化才刚起步，对于气象大数据的理解、应用和探索还在进行中，因此在气象大数据发展脉络把握，如何将大数据技术与气象业务相结合，如何开拓新的数据服务模式和数据服务创新机制等方面还有许多亟待解决的问题。

### 参考文献：

[1] 深圳国泰安教育技术股份有限公司大数据事业部群，中科院深圳先进技术研究院-国泰安金融大数据研究中心. 大数据导论—关键技术与行业应用最佳实践 [M]. 北京：清华大学出版社，2015：2-3.

[2] 李社宏. 大数据时代气象数据分析应用的新趋势 [J]. 陕西气象，2014（2）：41-44.