

全媒体人才培养丛书·数据科学系列

广播电视大数据 平台架构与应用

ARCHITECTURE AND APPLICATION OF
RADIO AND TV BIG DATA PLATFORM

林卫国 尚文倩 编著



中国传媒大学出版社

全媒体人才培养丛书·数据科学系列

广播电视大数据 平台架构与应用

ARCHITECTURE AND APPLICATION OF
RADIO AND TV BIG DATA PLATFORM

林卫国 尚文倩 编著



中国传媒大学出版社

·北京·

图书在版编目(CIP)数据

广播电视大数据平台架构与应用 / 林卫国, 尚文倩编著.
—北京: 中国传媒大学出版社, 2021.12
(全媒体人才培养丛书·数据科学系列)
ISBN 978-7-5657-3075-7

I. ①广… II. ①林… ②尚… III. ①广播工作—数
据处理—研究 ②电视工作—数据处理—研究 IV. ①G220

中国版本图书馆 CIP 数据核字(2021)第 209202 号

广播电视大数据平台架构与应用

GUANGBO DIANSHI DASHUJU PINGTAI JIAGOU YU YINGYONG

编 著 林卫国 尚文倩
策划编辑 阳金洲
责任编辑 黄松毅
封面设计 风得信设计·阿东
责任印制 李志鹏

出版发行 中国传媒大学出版社

社 址 北京市朝阳区定福庄东街 1 号 邮 编 100024
电 话 86-10-65450528 65450532 传 真 65779405
网 址 <http://cucp.cuc.edu.cn>
经 销 全国新华书店

印 刷 唐山玺诚印务有限公司
开 本 787mm × 1092mm 1/16
印 张 16.5
字 数 352 千字
版 次 2021 年 12 月第 1 版
印 次 2021 年 12 月第 1 次印刷

书 号 ISBN978-7-5657-3075-7/G · 3075 定 价 78.00 元

本社法律顾问: 北京李伟斌律师事务所 郭建平

版权所有 翻印必究 印装错误 负责调换



教授，博士生导师。现任中国传媒大学计算机与网络空间安全学院院长。

2011年博士毕业于中国传媒大学通信与信息系统专业，美国南加州大学(USC)电子工程系访问学者。目前主要研究方向是数字版权管理技术、大数据技术与人工智能在广播电视及融合媒体行业中的应用。作为第一作者和共同作者发表学术论文多篇，主持科研项目多项。曾获得广播电影电视部中青年优秀科技论文奖和国家广播电影电视总局高校优秀科研成果一等奖。承担过包括国家重点研发计划项目、国家科技支撑计划重点项目以及国家广播电影电视总局科技项目。



尚文倩:博士(后),教授,博导,现任中国传媒大学计算机与网络空间安全学院副院长。发表SCI/EI检索学术论文60余篇,论文单篇最高他引率300余次。曾主持或参与国家级、省部级、横向、校级等项目多项,荣获奖项多项,

IEEE/ACIS CSII2021 程序委员主席。

全媒体人才培养丛书·数据科学系列

大数据技术导论

数据可视化原理与实例

面向软件开发的实用数据可视化

社交网络分析

人工智能时代的新融媒体

图数据库的影视数据应用基础与示例

大数据时代的传媒信息安全

数据分析与数据挖掘实用教程

■ 广播电视大数据平台架构与应用

计算机网络协议分析与程序设计实践



服 务 传 媒 · 出 版 思 想

前 言

广播电视行业拥有海量的媒体内容数据、用户服务数据以及相关业务数据，如何充分挖掘这些数据的价值，为内容生产、业务创新、用户服务、运营推广、领导决策等提供支持，是一个全新的课题。国家广播电视总局已经将大数据科技创新与应用列为重点工作。

编写本书的目的是为各级广电网络运营商利用大数据技术、建设大数据平台提供指导，推动行业在大数据技术应用上开展创新实践，一方面可以促进广电企业各方面业务的发展，降本增效，充分利用大数据平台带来的技术优势，降低建设成本，提升运营质量和业务效益，增强行业竞争力；另一方面还能够更好地支持文化传播和信息服务，为广电大数据的发展提供参考和思路，注入新活力。

全书共6章，其中第1章是对大数据发展现状和趋势的概述；第2章在针对广电网络运营商的充分调研的基础上，研究和分析广电行业大数据的应用需求；第3章研究了广电网络运营商的数据资产范畴和数据资产管理思路；第4章围绕广电大数据应用体系建设总体目标研究符合企业实际要求的大数据应用体系建设规划；第5章从技术规划的角度阐述广电网络运营商大数据平台的建设思路和总体架构，梳理大数据平台总体数据流程及层次结构，并概述基于该平台的大数据分析技术和实际应用场景。第6章研究并讨论广电网络大数据平台建设演进规划。

本书由尚文倩教授负责第1章、第2章、第3章的撰写，林卫国教授负责第4章、第5章、第6章的撰写。全书由林卫国教授负责统稿。参加本书编写的还有任思可、张春洁、宇文易男、海日汗、黄俊杰等。本书在编写过程中得到了广播电视同行的大力支持，在此一并表示感谢。

由于编者水平有限，此书与大家见面后，我们将虚心接受各位的批评与指正，并乐意和大家就广电大数据技术的应用问题展开探讨。

编者

目 录

- 第1章 大数据发展现状与趋势 /1

- 第2章 广播电视大数据发展现状与需求 /7
 - 2.1 广播电视大数据建设现状 /7
 - 2.1.1 数据来源 /7
 - 2.1.2 数据应用现状 /9
 - 2.1.3 数据标签体系建设现状 /11
 - 2.2 广播电视网络与IT平台现状及需求分析 /13
 - 2.2.1 IT支撑平台现状及需求分析 /13
 - 2.2.2 网络平台现状及需求分析 /27
 - 2.2.3 业务平台现状及需求分析 /28
 - 2.3 广播电视网络运营商各部门大数据需求分析 /29
 - 2.3.1 运维类需求 /30
 - 2.3.2 经营类需求 /32
 - 2.3.3 销售类需求 /34
 - 2.4 小结 /35

- 第3章 广播电视数据资产管理 /36
 - 3.1 数据资产管理理论及发展 /36
 - 3.2 广播电视行业数据资产管理现状 /37
 - 3.3 广播电视运营商数据资产范畴 /40
 - 3.3.1 广播电视网络数据资产概述 /41
 - 3.3.2 广播电视网络数据资产梳理 /44
 - 3.4 有线网络数据资产管理 /45
 - 3.4.1 数据资产管理必要性 /45

- 3.4.2 数据资产管理技术 /46
- 3.4.3 数据资产管理思路 /52
- 3.5 小结 /64

第4章 广播电视大数据应用体系建设规划 /66

- 4.1 大数据应用体系建设总体目标 /66
- 4.2 大数据应用体系建设规划思路 /66
- 4.3 大数据应用体系发展规划 /67
 - 4.3.1 战略愿景和目标 /67
 - 4.3.2 对内应用规划 /69
 - 4.3.3 对外经营规划 /83
- 4.4 大数据应用体系能力要求 /84
- 4.5 小结 /85

第5章 广播电视网络运营商大数据平台技术与规划 /86

- 5.1 大数据技术发展趋势 /86
 - 5.1.1 云计算与大数据 /86
 - 5.1.2 广播电视物联网 /88
 - 5.1.3 大数据技术研究方向与趋势 /90
- 5.2 大数据平台建设思路与技术规划 /101
 - 5.2.1 大数据平台总体架构 /101
 - 5.2.2 大数据技术研究需求分析 /109
 - 5.2.3 大数据平台总体数据流程 /110
 - 5.2.4 大数据平台层次结构 /112
 - 5.2.5 数据管理与安全 /164
 - 5.2.6 实现大数据平台弹性扩展 /169
 - 5.2.7 部门级大数据平台的整合与数据迁移 /180
 - 5.2.8 现有数据仓库向云计算大数据转化 /182
- 5.3 大数据分析与应用 /190
 - 5.3.1 大数据分析技术 /190
 - 5.3.2 大数据应用场景与技术 /193
 - 5.3.3 数据运营加速新媒体发展 /234
- 5.4 小结 /237

第6章 广播电视网络大数据平台建设演进规划 /239

- 6.1 大数据平台建设规划 /239

6.1.1	平台建设规划思路	/239
6.1.2	平台能力要求	/240
6.1.3	平台演进路径	/241
6.1.4	平台建设目标	/241
6.2	大数据平台演进规划	/243
6.3	小结	/249
参考文献		/251

第1章 大数据发展现状与趋势

经过多年信息化建设和互联网发展，中国已经成为数据大国，互联网已经成为生产、生活的重要技术平台。随着信息技术的日益普及和“互联网”融合态势的加快，海量数据不断快速聚集，据有关机构统计，中国的数据总量将以年均50%的速度增长，预计到2020年将占全球数据总量的21%。在新形势下，国务院于2015年6月24日印发了《关于运用大数据加强对市场主体服务和监管的若干意见》，提出要充分运用大数据先进的理念、技术和资源，加强对市场主体的服务和监管，推进简政放权和政府职能转变，提高政府治理能力。同年8月31日颁布了《关于促进大数据发展的行动纲要》，并提出“坚持创新驱动发展，加快大数据部署，深化大数据应用，已成为稳增长、促改革、调结构、惠民生和推动政府治理能力现代化的内在需要和必然选择”。

与此同时，全国各地也加快了推动大数据发展的步伐。其中，广东、上海、重庆、贵州、武汉等省市在全国率先制定颁布了推动大数据发展的效果政策文件。早在2012年11月，广东便发布了《广东省实施大数据战略工作方案》，提出“为保证大数据战略有效实施，广东省将建设政务数据中心，并为高等院校和企业等成立大数据研究机构提供支持”。2013年7月，上海发布《上海推进大数据研究与发展三年行动计划（2013—2015年）》，提出“突破大数据关键技术，建立以企业为主体、产学研联合的发展机制，形成需求牵引、创新应用的发展模式，发展数据产业，服务智慧城市”。拟成立“上海大数据产业技术创新战略联盟”，建设“上海市数据科学重点实验室”、数据工程技术研究中心等，以大数据技术创新及产业应用为目标、以联盟为纽带促进形成若干引领大数据产业技术创新的企业联合实体。2014年2月，贵州印发《关于加快大数据产业发展应用若干政策的意见》《贵州省大数据产业发展应用规划纲要（2014—2020年）》，提出贵州省大数据发展的目标是，推动贵州省大数据产业稳步快速发展，到2020年，大数据带动相关产业规模达到4500亿元。大数据产业体系基本健全，业务形态较为齐备，创新能力显著增强，安全保障能力明显提高。产业载体建设顺利推进，聚集一批具有较强市场竞争力的骨干企业。数据中心布局合理，政府数据资源实现有效整合，大数据开放与管理机制初步建立，应用水平明显提高，以大数据引领和支撑贵州省经济社会转型发展的能力显著增强。此外，重庆、武汉等其他省市也纷纷颁布

了各自的大数据行动计划。

党中央、国务院高瞻远瞩，高度重视我国大数据发展。2015年8月31日国务院发布了《促进大数据发展行动纲要》，指导我国大数据发展的国家顶层设计和总体部署。

随着新一代物联网、云计算、移动互联网等信息技术的不断普及，人类产生的数据量正在呈指数级增长，大约每两年翻一番，而且这个速度在2020年之前继续保持下去。这意味着人类在最近两年产生的数据量相当于之前产生的全部数据量。有关资料显示，2011年，全球数据规模为1.8ZB，可以填满575亿个32GB的iPad，这些iPad可以在中国修建两座长城。到2020年，全球数据将达到40ZB，如果把它们全部存入蓝光光盘，这些光盘和424艘尼米兹号航母重量相当。

在此情况下，数据日益成为一种重要的战略资源。近年来，美、欧、日等发达国家和地区也纷纷将开发利用大数据作为夺取新一轮竞争制高点的重要抓手，相继出台了相关的战略规划和配套法规来促进大数据应用与发展，致力于利用大数据来进一步增强自身的综合竞争力。

2010年，美国总统科学技术顾问委员会（PCAST）向奥巴马和国会提交的《规划数字化的未来：美国总统科学技术顾问委员会给总统和国会的报告》指出，大数据技术具有重要战略价值，而联邦政府对其研发投资不足，建议联邦政府“制定一个应对‘大数据’的战略”。此后，奥巴马政府着手制定了一系列主题为“我的大数据”（MyData Initiatives）的倡议和措施，分别为2010年发起的《“蓝纽扣”计划》、2012年发起的《“绿纽扣”计划》、2012年发起的《“我的学生数据”计划》和2014年发起的《“创建副本”计划》，旨在促进美国人可以更安全地获取他们的个人数据，从而更好地处理他们私人领域的申请活动和服务。2012年3月，美国投资2亿美元启动了“大数据研究和发展”计划，该计划涉及国防部、能源部等6个联邦政府部门，用以大力推进大数据的收集、访问、组织和开发利用等相关技术的发展，进而大幅提高从海量复杂的数据中提炼信息和获取知识的能力与水平。政府对大数据带来的负面影响也更加重视，白宫2014年5月发布的白皮书《大数据：抓住机遇，守护价值》提出：“大数据正在改变世界。但是它并没有改变美国人对于保护个人隐私、确保公平或是防止歧视的坚定信仰。”

欧盟于2011年12月公开颁布了《开放数据战略》。它以开放数据为核心，以创新、增长和透明治理为引擎，重点加强在数据处理技术、数据门户网站和科研数据基础设施三方面的投入，旨在使欧洲企业与市民能自由获取欧盟公共管理部门的所有信息，建立一个汇集不同成员国以及欧洲机构数据的“泛欧门户”。目前，欧盟正在研究制定数据价值链战略计划，以实现数据的最大价值，重点是通过一个以数据为核心的连贯性欧盟生态体系，让数据价值链的不同阶段产生价值。数据价值链战略计划包括开放数据、云计算、高性能计算和科学知识开放获取这四大战略，遵循的主要原则是：高质量数据的广泛获得性，包括公共资助数据的免费获得；作为数字化单一市场一部分，欧盟内数据的自由流动；寻求个人潜在隐私问题与其数据再利用潜力之间的适当平衡，

同时赋予公民以其希望的形式使用自己数据的权利。

2012年6月,日本IT战略本部发布电子政务开放数据战略草案,迈出了政府数据公开的关键性一步。为了确保国民方便地获得行政信息,政府将利用信息公开方式标准化技术实现在紧急情况下能以较少的网络流量向手机用户提供统计信息、测量信息、灾害信息等公共信息,并尽快在网络上实现行政信息全部公开并可被重复使用的目标。2012年7月,日本总务省ICT基本战略委员会发布的《面向2020年的ICT综合战略(草案)》提出,“提升日本竞争力,大数据应用不可或缺”,这一战略将重点关注“大数据应用”。2013年6月,安倍内阁正式公布了新IT战略——《创建最尖端IT国家宣言》,全面阐述了2013—2020年期间以发展开放公共数据和大数据为核心的日本新IT国家战略,提出要把日本建设成为一个具有“世界最高水准的广泛运用信息产业技术的社会”。

在借鉴国外发达国家和地区先进经验的基础上,充分考虑中国具体国情及国内外大数据发展的趋势,我国促进大数据发展的策略包括如下几个方面:

1. 推动数据开放

政府拥有大量有价值的信息,因此政府应该着眼于民生福祉,带头推动政府数据的开放共享。要加快建立政府数据开放平台,推动各级政府和公共服务机构优先开放高价值数据资源。加强政府与民间协作,推动各类优质数据资源的开放共享。积极发挥市场的决定性作用,引导更多非公共数据向社会开放,鼓励基于开放数据开展应用创新,通过数据资源的开放利用,促进大数据技术和产业创新发展,并惠及更多民众。

2. 健全法律法规

借鉴美欧等发达国家和地区的经验,加快制定完善《中华人民共和国政府信息公开条例》《中华人民共和国保守国家秘密法》《中华人民共和国个人信息保护法》等有关信息公开、数据安全、网络安全、隐私保护方面的法律法规,规范国家重要数据的存储、备份和迁移,保障数据的安全可靠。在尊重知识产权的前提下,规范数据隐私保护的范 围、方式和程序,有效避免信息公开与隐私保护方面的冲突,对相关数据去除敏感信息后实现开放和共享。

3. 完善标准规范

尽快制定并试行有关数据开放原则和机制规范、数据分级标准、数据发展及使用的责任与权益等大数据发展应用过程中必需的相关标准规范。制定数据编码、处理、共享、交换标准;出台大数据技术、协议、标准等规范;建立数据资产访问、连接和共享机制,搭建数据资产交易平台,形成数据流转的层次化体系结构;研究数据资产的所有权、使用权以及价值评估体系,通过市场化模式保障数据资产流转的可行性。

4. 保障信息安全

妥善处理发展创新和安全规范的关系,探索完善安全管理规范措施,切实保障数据安全。建立大数据安全评估体系,做好大数据应用安全评测和风险评估,提高大数据平台信息安全监测、预警和应对能力。加强测试工具研发,开展大数据平台可靠性及安全性评测服务,引导大数据安全可控和有序发展。推动数据保护、个人隐私、数

据权益和合理利用等方面的地方立法工作，加快出台实施细则，对重要数据的保存、备份、迁移等进行规范管理。

5. 加强人才培养

落实各项人才政策，创建吸引人才、留住人才的良好环境，加快从美、欧等发达国家和地区引进既懂业务知识又具分析技能的复合型数据人才。扶持高等院校大数据相关专业的专业发展，推进大学课程改革，开设一系列符合现在和未来社会需求的数据技术的相关课程，培养数据存储、数据挖掘、数据可视化等方面的专门人才。鼓励高校和企业通过建立联合实验室、研发中心等形式，联合培养理论与实践相结合的大数据专业人才。以岗位培训和继续教育为重点，加大社会化培训力度，提高在职人员大数据应用技能，进一步加强对机关公务员、事业单位、大中型企业工作人员的大数据知识的培训与考核。

6. 支持技术创新

加大财政对大数据关键技术研发创新的引导和扶持力度，构建自主的大数据技术体系。加强大数据共性基础技术研发，重点研发大数据存储管理、分析挖掘、可视化、数据安全等关键技术和产品，提前布局自然语言处理、语义理解和机器学习等关键技术。支持国内创新型企业，开发专业化的数据处理分析技术和工具，提供特色化的数据服务。支持高校和科研院所建立大数据开放实验室，推动多学科交叉融合，开展大数据分析关键算法、关键技术研究 and 大数据专业人才培养，并加强实践应用。

广播电视行业拥有丰富的数据资源，特别是随着近年来的数字化、网络化和媒体融合发展，内容生产由模拟信号处理方式向全流程的数字化、网络化转变，传输分发由单向发射传输向双向数字网络转变，内容消费方式由简单接收向多元化互动方式转变，广播电视行业在内容生产、传输分发和客户服务过程中产生的数据规模快速增长，数据体量十分庞大。国家广播电视总局《2018年全国广播电视行业统计公报》公布：2018年全国广播节目制作时间为801.76万小时，全国电视节目制作时间为357.74万小时。2018年全国公共广播节目播出时间为1526.74万小时；全国公共电视节目播出时间为1925.03万小时。2018年全国有线广播电视实际用户数为2.18亿户，全国有线数字电视实际用户数为2.01亿户。2018年全国交互式网络电视（IPTV）用户为1.54亿户，互联网电视（OTT）用户为4.20亿户。广电行业海量的音视频文件及元数据蕴藏着巨大的价值，亟须引入大数据的思维和技术。第一，大数据可以为内容创作者和运营者的决策提供数据支撑，提升用户黏性以及业务收入；第二，大数据可以为行业提供宏观统计数据，促进产业发展；第三，大数据可以为政府提供产业及舆情信息，保障文化产业健康发展。

国家新闻出版广电总局科技司于2014年7月发布“基于大数据的网络广播电视台用户行为分析及推荐方法研究”科研项目，由广播科学研究院牵头，央视、央广、江苏总台共同参与，开展基于IPTV、网络电视台、网络广播台收视纪录的大数据研究工作。新疆广电网络开展广电网络大数据研究应用工作，建设了“天山云大数据指南针

研究室”，从地域、语言、业务等多维度，对全网数十万双向用户进行数据采集和分析，并每周定期公布“天山云指数”。广东广电网络股份公司在东莞投资建设了“广东广电大数据产业中心”大型文化产业项目，项目已入选广东省政府《“宽带广东”5年发展规划（2014-2020）》。在舆情分析方面，湖南台强档节目《爸爸去哪儿》是利用大数据手段研究电视舆情的典型应用案例。2013年，《爸爸去哪儿》播出的同时，数据分析团队通过抓取新浪微博上提及“爸爸去哪儿”的45.5万条原创微博数据，对36.7万独立原创作者用户、1300余万条用户微博及近1亿的关系进行数据分析，将《爸爸去哪儿》的美誉度与《中国好声音》《快乐男声》《天天向上》《快乐大本营》《非诚勿扰》等节目进行了舆情数据对比分析，《爸爸去哪儿》成为名副其实的“口碑王”。舆情数据还对该档节目的粉丝性别、地域、年龄、媒体偏好度进行了统计分析，对于节目中的主角和笑点事件都做了统计分析，各位爸爸的热度变化趋势一目了然。最终还对三大赞助商的微博热值给出了参考值。此外，各地局、台、网也相继开展了大数据的研究与建设，利用大数据提升运营质量。

广电“十三五”规划对广电行业提出了如下发展要求，要显著提升广电终端的智能化水平，提高安全播出和信息安全能力，推动云平台建设，为建设智慧社区、智能家庭和智慧城市等做好铺垫。

在《新闻出版广播影视“十三五”科技发展规划》发布的新形势和大背景下，国家新闻出版广电总局对于广电大数据的建设发展提出了一系列重要的目标和任务，下面列出了其提出的部分重要发展目标和主要任务。

发展目标：显著提升广电终端标准化、智能化应用能力。大力推进TVOS在有线、无线、卫星和互联网等广电终端的应用，基本实现机顶盒、一体机、媒体网关和智能终端等广电终端的标准化、智能化，形成广播电视和互联网融合的“广电+”生态，广电智能终端成为智能家庭、智慧社区和智慧城市的重要基础。主要任务包括：

（1）大力推进广播电视融合媒体服务云建设，打造“广电+”生态链，推动广播电视系统服务的升级转型；

（2）积极推动广播电视技术与新一代信息技术融合创新，全面增强广播电视科技创新能力。

除此之外，国家新闻出版广电总局还提出了一些和广电大数据相关的重点项目：

（1）广播电视融合媒体制播大数据关键技术研究与应用示范；

（2）广播电视台融合媒体用户互动数据采集接入技术研究；

（3）融合媒体服务与传输大数据关键技术研究与应用示范。

在《新闻出版广播影视“十三五”科技发展规划》中，显著提升广电终端标准化、智能化应用能力已被列为重要发展目标，明确提出要使传统广电行业和新型信息技术进一步融合，广电面临全面的战略转型，广电网络运营商更加面临着前所未有的挑战与机遇。

《新闻出版广播影视“十三五”科技发展规划》提出，广电要基于TVOS技术，不

断推进广电融合发展，实现有线、无线、直播卫星电视、融合终端等广电终端基本标准化、智能化，让广电拥抱互联网，实现融合媒体服务云平台的建设。要运用云计算、大数据技术，实现广电各方面资源的有效整合，最终使其得以共享和开放，让更多的人受益，为开展更多的广电业务提供基础支撑。既能让融合媒体云平台在不同的系统下运行，也能够使平台上的应用运行在云端与终端随意地切换，为用户带来一体化的体验。

具体到广播电视网络运营商，他们目前也面临着行业竞争的压力和发挥数据资产潜力的需求，在当下各行各业都广泛应用大数据的背景下，通过大数据平台解决公司内外部需求和矛盾显得尤为重要。目前广播电视网络运营商对于大数据的需求可以从多个维度来看待，目前各广播电视网络运营商已经建设起自己的数据仓库，拥有很多高质量的数据资产，但是长期的烟囱式系统架构带来的数据流通不畅、部门协调差、数据利用率不高等问题日渐突出，给企业运营和运维都带来了很大的不便。总体来看，广播电视网络运营商对于大数据的需求可以分为以下几类：运维类需求、经营类需求，以及销售类需求。运维类需求在于改变广播电视网络公司传统运维现状，提高运维工作效率，在设备出现问题的时候，可以利用大数据工具，采集，分析数据，精准地找到设备的故障点，在用户反馈之前发现故障，解决问题，最终达到提高用户体验的目的；经营类需求在于明确用户需求，利用大数据技术进行精准推荐，对应用类和视频类的引进和推广；销售类需求在于通过大数据平台建设构建以客户为中心的市场营销、客户服务和营销管理体系。当下广电行业形势严峻，穷则思变，互联网时代需要摒弃门户之见，改变思想，学会利用互联网技术来发展传统企业，当下OTT和IPTV的用户量的剧增，OTT和在线视频网站对广电传统业务产生了巨大的冲击，但是广电拥有自己的优势，如用户体量大，数据价值高，基础管道建设良好等。在面临挑战的同时，利用好自身优势，抓住机会发展自己，明确自身需求，找准发展方向，就一定能够逆势而上，创造新的局面。

第2章 广播电视大数据发展现状与需求

2.1 广播电视大数据建设现状

广播电视行业作为拥有用户量最多的行业之一，有着上亿用户的支持，广播电视台每日也会不断产生大量的节目数据、用户数据和用户与平台互动的数据等。这些数据积少成多，规模较大，需要的存储空间也不断加大，因此对数据处理速度也提出了较高的要求。广播电视行业面临的严峻形势不可小觑，广播电视行业需要抓住机遇与挑战，不断丰富自我，发挥自身优势，在这个互联网时代，创造更加辉煌的业绩。

广播电视行业的企业了解自身的状况对发展自我是至关重要的，下面我们将通过分析典型广播电视网络运营商的数据来源、数据的应用现状、数据仓库建设现状、数据标签体系建设现状以及各部门的需求等方面，力求在明确自身现状的基础上，规划出一个有价值的大数据建设发展方案，供广电企业进行参考。

2.1.1 数据来源

广播电视行业收集的数据类型多样，经过数年的发展，已经形成了一定规模的数据体量。由于三网融合的不断推进，广播电视行业不断受到来自互联网的影响，相对于以往便于存储的以元数据和媒体视音频数据为主的结构化数据，来自于互联网的非结构化数据大量涌入，包括网络文字、音频、视频、图片、地理位置信息等；同时节目的全媒体融合制作产生了丰富的内容资源；互动数据的接入使节目播出形式多样化；对用户数据的收集也不再拘泥于用户的基本信息，用户的收视行为也一样成为了收集和分析的重点。基于这些情况，大体可将广播电视网络公司的数据分为结构化、半结构化以及非结构化数据，具体细节如图2.1所示。

广播电视网络运营商现阶段的数据来源包括该企业客户的资料、为客户提供订购服务产生的数据、终端设备产生的数据、用户的收视行为数据以及上网行为数据、运维网络上各个网点的日志数据、工作人员对节目内容进行编排产生的数据。

图2.2为某广播电视网络运营商近十年来的数据资源收集变化情况。

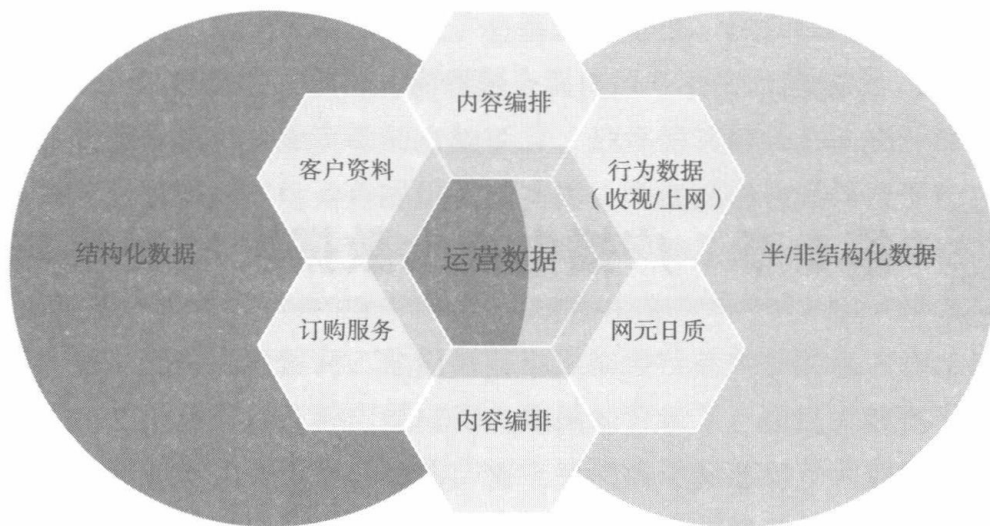


图 2.1 公司数据资源概况

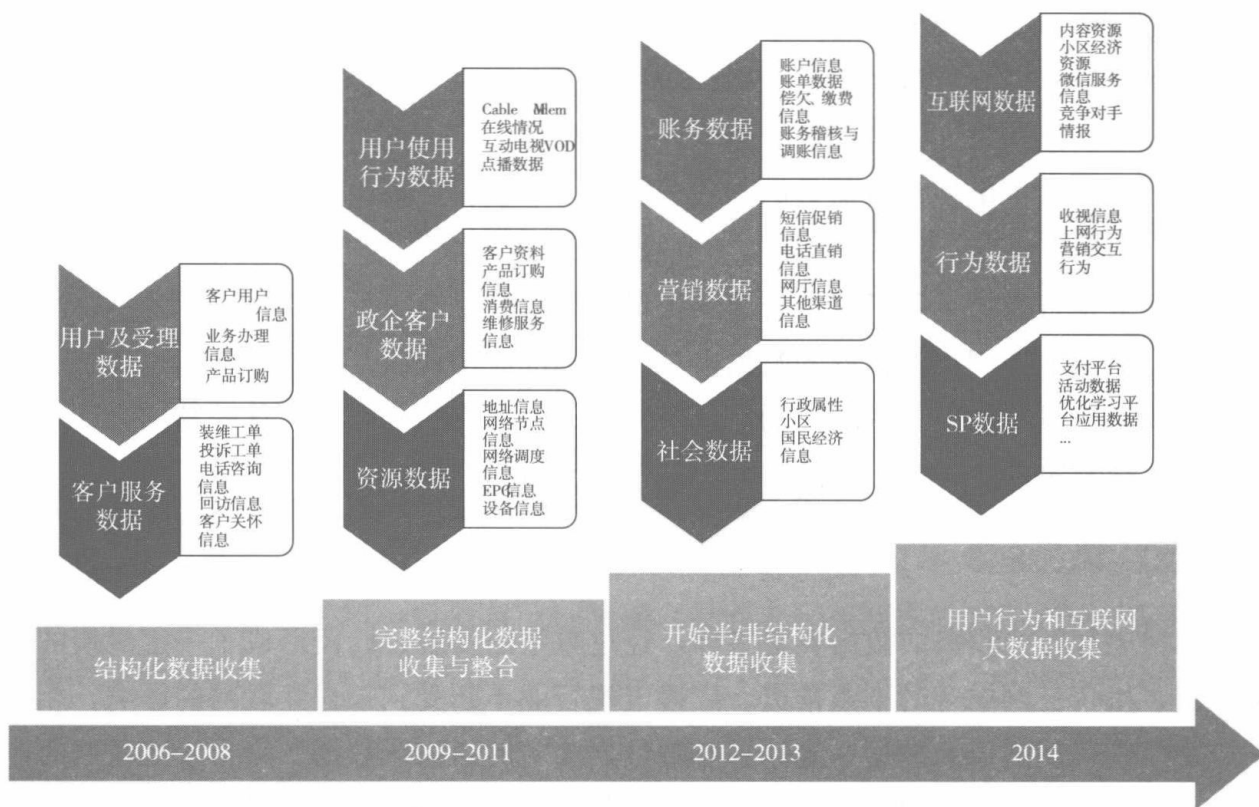


图 2.2 某广播电视网络运营商数据资源概况

2006年至2008年，该运营商主要收集一些结构化的数据。这些数据来源仅限于用户信息以及为其服务所产生的各类数据，其中包括客户用户信息、业务办理信息、产品订购信息、为维修用户的终端设备所产生的运维工单、受理用户投诉信息的投诉工单、用户电话咨询业务产生的相关数据。

2009年至2011年，该运营商完成了结构化数据的收集以及整合，除上述2006年至2008年收集的数据以外，该运营商开始收集用户使用终端设备产生的行为数据、政企客户的数据信息，以及各种设备资源产生的信息。其中用户使用设备产生的行为数据，包括用户网络在线使用行为数据以及用户点播电视节目产生的行为数据；政企客户数据包