

航空 大数据

聚数据之力，启航空之智

管涛◎著



 中国工信出版集团

 电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

航空大数据

管涛著

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书围绕若干航空领域的业务，分析了该领域面临的大数据需求，指出大数据分析的目的、数据来源和存在的困难，给出了数据分析的方法和大数据平台参考方案。书中内容分为七章。第一章介绍了大数据的背景与概念；第二章简述了一些相关的计算方法，演示了一些算法的使用方法和过程；第三章介绍了工业大数据相关的概念，探讨了航空制造大数据的业务模式及在生产中的作用，提出了基于制造执行系统（MES）的工业大数据分析模型。第四章描述了航空物流大数据概念，介绍了航空物流的业务模式；第五章分析了跨境电商的主要业务，提出了面向跨境电商的大数据平台及各个业务系统；第六章介绍了航空客流的主要业务，列举了两个典型的案例。第七章介绍了面向企业的大数据服务。附录给出了中英文词汇对照表、国内外大数据政策、著名的数据源及参考文献。

本书适于航空信息、计算机、电子信息、信息管理、档案学等专业的师生阅读，亦可作为航空工业相关科技人员、大数据工程技术人员、教育培训的参考用书。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目（CIP）数据

航空大数据 / 管涛著. —北京：电子工业出版社，2018.6

ISBN 978-7-121-33717-8

I. ①航… II. ①管… III. ①数据处理—应用—航空运输管理 IV. ①F560.8-39

中国版本图书馆 CIP 数据核字（2018）第 031213 号

策划编辑：祁玉芹

责任编辑：鄂卫华

印 刷：中国电影出版社印刷厂

装 订：中国电影出版社印刷厂

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：787×1092 1/16 印张：9 字数：202 千字

版 次：2018 年 6 月第 1 版

印 次：2018 年 6 月第 1 次印刷

定 价：39.80 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zllts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）68253127

前 言

preface

大数据是一个时代的主题，有力地推动了社会生产力的发展，促进生产关系高度演化、高效协同，生产资料优化配置，推动着劳动者从人向机器的转变。目前，大数据已经深入到工程应用的各个领域，如电商、医疗、农业、交通、金融、安全、侦查、城市治理、天气预报，并带来了显著的效益和价值。在航空工业，人们经常采集和使用大量的航空领域数据，用来监测、评估、诊断、模拟、仿真、设计和决策，但是，完整地、系统地阐述和使用航空大数据仍是一个未知的领域。有效利用航空大数据，挖掘内在的价值，为航空产业服务已经成为航空领域亟待发展的方向和研究的课题。然而，由于航空领域涉及的业务领域众多，很难用一个统一的框架进行描述。同时，每个领域都有独特的业务流程，增加了大数据分析的难度。

航空大数据是全新的概念，与航空生产、贸易、物流、产品、服务、信息化有关，专业性强，涉及范围广泛。本书从航空制造、航空物流、跨境电商、航空客运、企业服务五个领域入手，期待理清每个领域的业务流程，探讨航空大数据的来源、分布、特点、价值和利用，形成较为完整的大数据分析思路，构建面向领域的解决框架，给出航空大数据分析和挖掘的参考模型、方法和解决方案，从而为后续的工程实践奠定基础。通过一些案例分析，作者解释了大数据应用的方法和价值。本书理论结合实践，采用图文并茂的形式、简洁的语言阐述基本概念、技术、航空大数据的业务和研发思路，但是不包含具体的数学建模和软件开发、程序语言和代码，因为应用细节过于繁琐，不具有代表性。

自 2000 年以来，作者一直从事数据分析和挖掘相关领域的研究和教学工作。还在数学系学习的时候，接触的第一种分类模型即是支撑向量机，那时是从数学优化的角度考虑该问题。后来在计算机学院接触到韩家炜等人撰写的英文版《数据挖掘》，感受到数学模型和应用结合的重要性。岁月流逝，潮流变迁，大数据、人工智能逐步成为时代的热点。应用的快速发展解决了很多实际问题，然而，万变不离其宗，大数据分析的数学基础仍然是传统数学体系，很少发生变化。从应用场景抽象出来的大数据问题具有综合性、跨领域的特点，解决此类问题往往需要具备完善的、多方向的数学基础知识。因此，很好地分析大数

据的价值、构思大数据挖掘模型，就需要具有完善的数学基础知识积累。与此同时，大数据亦具备完善的工程体系，从数据采集物联网平台到分布式系统、系统框架、系统工具，都要求开发者具有完善的计算机理论基础和丰富的实践经验。大数据的应用模型与行业紧密结合，是行业分析的基础数据工具，具有跨领域特点。熟悉领域背景知识、了解行业需求、解决行业关键问题是应用的核心。

航空大数据的基础研究、产业及应用将随着新技术的出现而发展，与物联网、人工智能、云计算、移动计算、机器人、视频分析、GIS、智慧城市、共享经济的深度融合亦将成为未来的必然趋势，由此带给人们更智能、便捷、流畅的服务。顺应科技发展潮流、未雨绸缪，则一定会在未来的航空大数据深度发展中占据一席之地。

本书得到了郑州航空工业管理学院信息科学学院刘永教授的一些指导和建议，及滕立博士的热心帮助。同时，信息科学学院的研究生庞宇飞、任照博、任妍积极地参与了本书终稿的修改和校正工作，感谢他们付出的辛勤劳动。本书的出版得到了河南省信息智能处理工程技术研究中心、航空经济发展河南省协同创新中心的资助和支持。同时，也得到了河南省高等教育教学改革研究与实践项目（2017SJGLX406）、河南省重点研发与推广专项支持项目（182102210454）及郑州市普通科技攻关计划项目（20130783）的支持，在此一并表示感谢，同时感谢家人的大力支持。

管涛

航空经济发展河南省协同创新中心

2017年4月5日于郑州

目 录

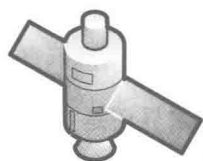
contents

第 1 章 引言	1
1.1 大数据概念	2
1.2 大数据的知识体系	4
1.3 数据思维	5
1.4 物联网	6
1.5 云计算	7
1.5.1 概念	7
1.5.2 架构	7
1.6 人工智能	9
1.7 大数据产业链	10
1.7.1 数据采集	10
1.7.2 数据清洗	11
1.7.3 数据存储	12
1.7.4 数据分析和挖掘	12
1.7.5 数据可视化	12
1.8 大数据工程的实施	13
1.9 大数据平台工具简介	14
1.9.1 Hadoop	14
1.9.2 Spark	17
1.9.3 Storm	18
1.9.4 Sqoop	18
1.9.5 Zookeeper	18
1.9.6 Kafka	19
1.9.7 数据仓库	19

1.10	开源大数据挖掘工具简介	22
1.10.1	Mahout	22
1.10.2	Weka	23
1.10.3	Pentaho	23
1.11	航空领域应用与发展	23
	小结	24
第2章 数据分析技术		25
2.1	引言	26
2.2	基本数据分析方法	26
2.2.1	信号处理方法	26
2.2.2	聚类与分类	28
2.2.3	关联分析	31
2.2.4	预测分析	35
2.2.5	层次分析法 (AHP)	38
2.2.6	深度学习	40
2.2.7	强化学习	41
	小结	42
第3章 航空制造大数据		43
3.1	简介	44
3.1.1	工业大数据	44
3.1.2	工业云	45
3.1.3	信息物理系统 (CPS)	46
3.1.4	制造执行系统 (MES)	46
3.1.5	大数据驱动的智能制造	47
3.2	航空制造业务模式分析	52
3.2.1	优化产品设计	53
3.2.2	智能车间	54
3.2.3	智能工厂	55
3.2.4	在线运维	57
3.3	基于工业流程管理的数据分析	57
3.3.1	系统框架	57
3.3.2	数据分析框架	60
	小结	62

第 4 章 航空物流大数据	63
4.1 航空物流概况	64
4.2 航空物流业务分析	66
4.2.1 仓储	66
4.2.2 航空货运	67
4.2.3 物品配送	68
4.2.4 海关监管	70
4.2.5 产品维护	74
4.3 数据分析系统	74
4.3.1 大数据分析与管理平台	74
4.3.2 综合展示和分析平台	76
4.3.3 多式联运数据分析	76
4.3.4 物流管理与决策分析	77
4.3.5 托运人及货代管理	78
小结	79
第 5 章 跨境电商大数据	81
5.1 跨境电商发展现状	82
5.1.1 宏观政策	82
5.1.2 存在不足	83
5.1.3 新商业系统	83
5.2 跨境电商业务分析	84
5.2.1 跨境仓储	85
5.2.2 支付体系	86
5.2.3 安全体系	86
5.2.4 质量体系（追溯体系）	87
5.2.5 统计体系	89
5.2.6 商品体系	89
5.2.7 物流成本	89
5.3 数据来源与存在的问题	90
5.4 跨境电商大数据参考平台	91
5.4.1 应用系统功能	92
5.4.2 数据流和综合业务展示	92
5.4.3 数据交换、共享和汇总	93

5.4.4	数据仓库	93
5.4.5	商品推荐系统	95
	小结	97
第 6 章	航空客运大数据	99
6.1	发展现状与政策	100
6.2	航空客运业务分析	101
6.2.1	机场管理与运营	102
6.2.2	航班的动态信息	103
6.2.3	旅客服务	103
6.2.4	航线管理与优化	104
6.2.5	燃油管理	105
6.3	大数据分析模型	105
6.3.1	数据仓库	106
6.3.2	企业效益分析	110
6.3.3	旅客画像	110
6.3.4	风险管理	112
6.3.5	供应链管理	113
6.4	案例分析	114
	小结	114
第 7 章	面向企业的大数据服务	115
7.1	引言	116
7.2	企业征信	116
7.3	企业知识图谱	120
7.4	企业专利地图	125
	小结	126
附录 A	中英文对照表	127
附录 B	国外大数据政策	130
附录 C	国家大数据和人工智能政策	131
附录 D	测试数据集	133
	参考文献	134



第 1 章 引言

- ❖ 1.1 大数据概念
- ❖ 1.2 大数据的知识体系
- ❖ 1.3 数据思维
- ❖ 1.4 物联网
- ❖ 1.5 云计算
- ❖ 1.6 人工智能
- ❖ 1.7 大数据产业链
- ❖ 1.8 大数据工程的实施
- ❖ 1.9 大数据平台工具简介
- ❖ 1.10 开源大数据挖掘工具简介
- ❖ 1.11 航空领域应用与发展

1.1 大数据概念

大数据与矿产资源、淡水、煤炭一样，是一种战略资源。无论是学术界还是企业界，都给予了大数据足够的重视，在研究和应用领域投入了大量的资金、人力和物力。2008年，《Nature》出版 Big Data 专刊探讨大数据的定义、形式和用途。2011年6月，麦肯锡发布了大数据报告 *Big data: The next frontier for innovation, competition, and productivity*，详尽分析了大数据的影响、关键技术和应用。2012年1月份，达沃斯世界经济论坛发布了报告 *Big data, big impact: New possibilities for international development*，探讨了如何更好地利用数据来产生良好的社会效益。国际数据公司（IDC）的研究报告称，2011年全球被创建和被复制的数据总量为1.8ZB，并预测到2020年，全球将拥有35ZB的数据量。大量的数据记录了人们参与社会活动的信息，必然蕴含人类生产活动的规律。

近些年来，国内出台了許多促进大数据发展的政策。2015年6月，《国务院办公厅关于运用大数据加强对市场主体服务和监管的若干意见》指出，要充分运用大数据先进理念、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。2016年1月，工信部印发了《大数据产业发展规划（2016—2020年）》，“提出了发展目标，将酝酿开启万亿级别市场规模，到2020年，大数据相关产品和服务业务收入突破1万亿元，年均复合增长率保持30%左右；将建设10~15个大数据综合试验区，创建一批大数据产业集聚区，形成若干大数据新型工业化产业示范基地”。2016年9月5日，国务院日前印发《促进大数据发展行动纲要》，系统部署大数据发展工作。2016年9月，国家重点研发计划启动实施“云计算和大数据”重点专项。2016年10月8日，国家发展改革委、工业和信息化部、中央网信办发函批复，同意在京津冀等七个区域推进国家大数据综合试验区建设，分别是两个跨区域类综合试验区（京津冀地区、珠江三角洲地区）、四个区域示范类综合试验区（上海市、河南省、重庆市、沈阳市）、一个大数据基础设施统筹发展类综合试验区（内蒙古）。2017年12月8日，中共中央政治局就实施国家大数据战略进行第二次集体学习。习近平主席提出要推动大数据技术产业创新发展，构建以数据为关键要素的数字经济，运用大数据提升国家治理现代化水平、促进保障和改善民生，切实保障国家数据安全。

大数据（Big Data）概念伴随着信息化、数字化技术的广泛普及和快速发展而出现，记载着人们工作、学习、娱乐过程中的每个细节和变化，在航空制造、智能电器、文本检索、电子商务、农业生产、社会治理、城市交通、智能车间、智慧政务、快递物流等领域快速发展和深度应用，带来了良好的服务和可观的收益。例如，定制化制造满足了人们个性化的需求；精准农业节约了大量的淡水、化肥资源；航空发动机通过传感器收集状态数

据，指导机器正常运转和远程维护；人们通过 12306 网站很方便查询和订阅火车票，避免了排队买票的尴尬；京东、淘宝电商推荐系统为人们提供最新的商品咨询，极大减少了查找的烦恼；百度地图导航解决了自驾出行的路线问题，方便了家庭的外出游玩；微信支付电费、水费、天然气费极大地方便了市民的生活；滴滴打车解决了人们出门“打车难”的问题；网格化智慧政务方便了社区管理、社区服务和城管信息上报，及时指导管理部门决策。

从狭义上看，大数据指数据本身，具有四个显著的特点：大容量、时效性、多样性、有价值，如图 1.1 所示。大容量指数据量大、维度高，占用大量的物理存储空间，普通计算机难以完成承载存储和简单的分析任务。时效性指数据实时更新、时间和数值精度要求高，比如日志数据、工控数据、监控数据，传输强度大、时效性高。多样性指数据的异构性强、种类繁多，如文本、数值、符号、多媒体、序列、对象、结构、GIS 和自定义格式等。有价值指大多数数据集合中蕴含了对象的有价值的知识和信息，虽然在许多情况下价值密度相对较低。从广义上理解，大数据是一个体系，包括数据和相关要素、工具，如智能算法、开源软件、硬件平台、应用场景。大数据离不开智能技术和分析过程。大数据是资源，智能分析是工具和手段，二者的有效协同产生了知识。开源软件简化了数据挖掘过程，节省了大量的人力。硬件平台是大数据软件开发的基础环境。应用场景是大数据的来源和实践之地。

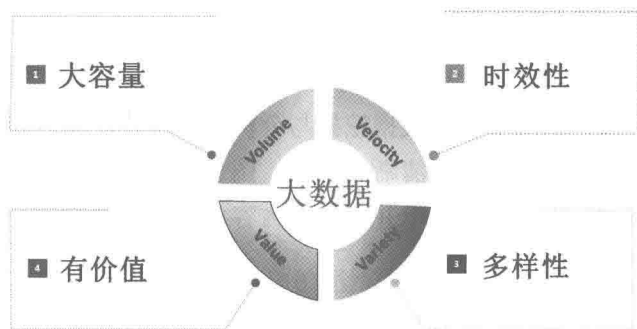


图 1.1 大数据的四个特点

在许多应用场景中，大数据与云计算架构和物联网平台紧密相关，如图 1.2 所示。人们通过物联网中各类传感器、移动终端等收集对象信息，然后通过网络传输到云端分类存储，日积月累形成各种类型的大数据。大数据在云端预处理、分析和挖掘，发现有价值的信息，反馈给决策者或者用户。当然，大数据也可以不存储在云端，而是存放在自建的大数据中心，选择何种方式取决于应用需求、运营成本和用户设计。

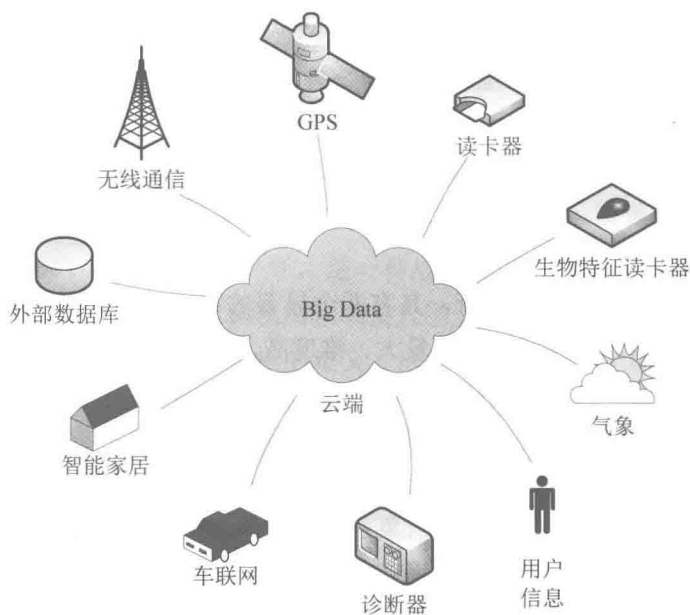


图 1.2 大数据与云计算、物联网的关系

1.2 大数据的知识体系

大数据包括了一个从理论到实践、多学科交叉的知识体系，涵盖了应用数学基础理论、计算机基础理论、大数据系统架构、数据挖掘与分析、开源软件平台和工具、大数据应用案例几大领域。为了从宏观上有一个清晰的认识，图 1.3 总结了大数据相关的知识体系。大数据涉及许多课程的若干知识点，但并非所有内容，因此，在学习、培训和实践的时候应当有所取舍。



图 1.3 大数据的基本知识体系

大数据不断与新兴技术、新兴模式交叉、融合，如物联网、智能制造、脑科学、共享经济、生命科学、下一代通信网络、IPV6、移动计算、北斗定位系统、GIS、虚拟现实，如图 1.4 所示。诞生了崭新的前沿领域和应用，深刻推动着产业的发展，丰富着人们的生活。



图 1.4 大数据与其他学科的交叉融合

1.3 数据思维

逻辑思维和计算思维是人们耳熟能详的思维方式。前者注重问题的推理和演算过程，在形式化问题的证明和求解中广泛应用。随着计算机的广泛普及，很多手工难以完成的任务都可以通过计算机实现。如此，把计算机看作一种工具，那么任何一种软件应用都可以看作一种广义的计算行为，计算思维由此产生。随着信息技术的进步，计算能力不再局限于计算机，任何拥有基本硬件结构的设备都具有计算和存储能力，如路由器、手机、洗衣机、电视机、智能插座等。如此多的计算设备给人们生活带来便利的同时，也带来了大量的数据。分析和处理这些数据、挖掘潜在价值是一大挑战。数据思维正是在这样的背景下产生的。数据思维过程是以数据为起点，以应用场景为目标，利用模型和方法产生价值的过程。与逻辑思维和计算思维比较，数据思维是一种唯物主义观，具有如下特点：客观性、抽象性、系统性、广泛性、可用性。

应用数据思维可以解决现实中的许多问题。以往的政府决策多依靠经验判断，通过协商的方式制定政策和规划。如今，通过管理信息系统，决策者可以获得大量的政策数据、

经济运行数据、各种反馈数据等，实现用数据说话，用数据决策。当前的传统企业多面临着互联网转型，本质就是企业经营过程的网络化、精准化、个性化、数据化。经营者通过数据及其关联即可发现企业生产流程情况、资金流转、销售情况、经营瓶颈等信息，针对性地开展决策和今后的规划。在日常生活中，地图导航解决了驾车外出旅游的尴尬；打车软件随时随地叫车，极大地方便了出行；共享单车解决了人们最后一公里的步行问题；公共服务软件免去了去服务站排队缴费过程，为大众缴纳燃气费、电费、水费和税费提供了便利。

政府提供服务的前提是了解大众的需求，通过调研、网络工具采集的大量数据为合理决策提供了参考依据，有效避免盲目、无效的决策。例如，在网格化城市管理中，网格员每天收集大量的社区信息，集中上报和管理，通过系统分析产生结论，为决策者提供参考，大大提升了社会治理的水平和层次。



1.4 物联网

国际电信联盟（ITU）发布的 ITU 互联网报告给出了物联网的定义：通过二维码识读设备、射频识别（RFID）装置、红外感应器、全球定位系统和激光扫描器等信息传感设备，按约定的协议，把任何物品与互联网相连接，进行信息交换和通信，以实现智能化识别、定位、跟踪、监控和管理的一种网络。

在大数据领域，物联网连接着物理世界，好像是中枢神经系统的感觉神经末梢，具有数据采集、网络传输、信息存储等功能。物联网链接物与物、人与物、人与人，从而构成了人、机、物为一体的信息空间网络。互联网将物联网的基础数据上载、集中到云端的大数据中心进行分析和处理。物联网有许多应用，如车联网、智能安防、物品识别定位、物流系统、生产设备状态监测、无线医疗监护、公交系统、智能停车、污水处理、农业生产等。当然，物联网在大数据体系中并非必选项，数据来源可以由其他途径获得。

物联网在许多领域有很好的应用。在精准农业领域，利用传感器监测植物不同生长阶段的二氧化碳、空气温度、土壤 pH 值、氮浓缩量、土壤水分、土壤湿度、光照强度、植物的病虫害。数据实时汇聚到中心控制系统，通过数据分析，得到植物在不同生长阶段的营养需求，启动相应设备进行调节，如开灯补充光照、喷水增加湿度，实现对作物生长的智能管理。在智能家居方面，通过无线物联网开展远程监护，实时收集老年人或者残疾人的心跳、血压、睡眠状态、起居情况等，把电子信息记录上载、分析，得到实时健康状况，开展相应的医疗指导。在城市地下管网监控中，通过传感器收集各种管线的实时状态，实现可视化在线监控管网的运行，预防故障的发生，指导维修资源调配。在智能家居领域，可实现家用电器的远程控制、智能门锁、家居监控、智能机器人控制等功能。



1.5 云计算

1.5.1 概念

云计算是一种通过 Internet 以服务的方式提供动态可伸缩的虚拟化资源的计算模式。这种模式提供可用的、便捷的、按需的网络访问，进入可配置的计算资源（资源主要包括网络、服务器、应用软件、存储及服务）共享池。这些资源能够被快速提供，用户可根据个人或团体的需要对云计算的资源进行租赁^[1]。

虚拟化是云计算使用的核心技术之一，将有限的 IT 物理资源虚拟为多个逻辑系统。从用户角度来看，每个逻辑系统就像一个真实存在的、完整的 IT 物理系统，而真实的物理系统对用户透明。虚拟化的优势在于降低成本、资源整合、优化资源、高度灵活性、可扩展性和安全性。虚拟化类型分为操作系统虚拟化、应用程序虚拟化、桌面应用虚拟化等。

1.5.2 架构

云计算平台有三种基本的服务模式：IaaS、PaaS、SaaS^[1]。随着大数据应用的普及，出现了 DaaS 模式，即数据即服务。

1. IaaS

IaaS (Infrastructure-as-a-Service): 基础设施即服务。IaaS 通过网络向用户提供计算机（物理机和虚拟机）、存储空间、网络连接、负载均衡和防火墙等基本计算资源。用户在此基础上部署和运行各种软件，包括操作系统和应用程序。消费者通过 Internet 可以从完善的计算机基础设施获得服务。常用的 IaaS 平台如 OpenStack, Cloudstack, Rackspace 和 NASA 联手推出的云计算平台。

2. PaaS

PaaS (Platform-as-a-Service): 平台即服务。PaaS 实际上是指将软件研发的平台作为一种服务，以 SaaS 的模式提交给用户。平台通常包括操作系统、编程语言的运行环境、数据库和 Web 服务器，用户在此平台上部署和运行自己的应用。用户不能管理和控制底层的基础设施，只能控制自己部署的应用。

3. SaaS

SaaS (Software-as-a-Service): 软件即服务。它是一种通过 Internet 提供软件的模式，用户无需购买软件，而是向提供商租用 Web 软件来管理企业经营活动。SaaS 不仅减少了或取消了传统的软件授权费用，而且厂商将应用软件部署在统一的服务器上，免除了最终

用户的服务器硬件、网络安全设备和软件升级维护的支出，除了个人电脑和互联网连接之外，客户不需要其他 IT 投资就可以通过网络获得所需要软件和服务。云计算平台的基本架构如图 1.5 所示。

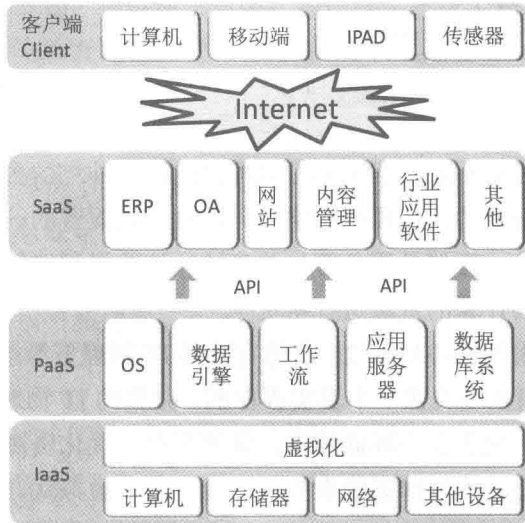


图 1.5 云计算架构示意图

随着大数据的发展，数据将成为一种服务，出现 Data-as-a-Service (DaaS) 模式。DaaS 通过数据的收集、整理、传输和分析，发现事物发展内部规律和模式，为用户提供便捷的商品销售指导、宏观市场分析、辅助决策、高品质生活等服务。例如，厂家通过用户购买数据了解市场需求，指导生产；政府通过宏观数据分析发现经济发展规律和存在的问题，制定相应的政策。DaaS 包括提供数据资源、数据存储、数据管理、数据共享与交换、数据分析、信息推送、数据展示等服务。

云计算平台和大数据处理之间关系如图 1.6 所示。

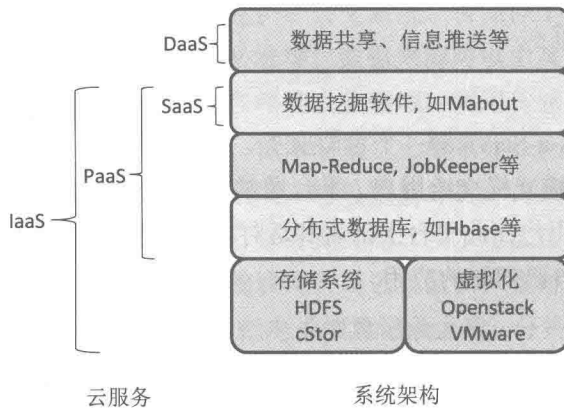


图 1.6 云计算和大数据部署关系