



物联网大数据

处理技术与实践

王桂玲 王强 赵卓峰 韩燕波◎编著



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>

物联网大数据处理技术与实践

王桂玲 王 强 赵卓峰 韩燕波 编著



电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书基于作者近几年来的研究开发成果及应用实践,对物联网大数据技术体系进行了系统归纳,阐述了物联网环境下感知数据的特性、数据模型、事务模型及调度处理方法等核心概念及关键技术,并对物联网大数据存储、管理、计算与分析的基本概念和关键技术进行了剖析。本书还介绍了自行研发的面向物联网的 ChinDB 实时感知数据库系统以及针对云计算环境下物联网大数据管理与应用的 DeCloud 云平台,介绍了它们在智能交通、智能电厂、教育、安全监控等多个行业的应用。书中所有实例,均来自作者所在团队的实际应用,大部分在物联网项目中得到了实践应用。本书对物联网应用的开发及两化融合、工业 4.0 环境下的大数据处理分析具有重要参考价值。

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

物联网大数据处理技术与实践 / 王桂玲等编著. —北京:电子工业出版社,2017.9

ISBN 978-7-121-32421-5

I. ①物… II. ①王… III. ①数据处理 IV. ①TP274

中国版本图书馆 CIP 数据核字(2017)第 189161 号

责任编辑:董亚峰

特约编辑:刘广钦 刘红涛

印 刷:三河市双峰印刷装订有限公司

装 订:三河市双峰印刷装订有限公司

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:720×1 000 1/16 印张:15.5 字数:299 千字

版 次:2017 年 9 月第 1 版

印 次:2017 年 9 月第 1 次印刷

定 价:49.00 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 zltz@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) 88254694。

前言



随着物联网产业的迅速发展，物联网应用也正由离散的、示范性应用逐步走向多层次、规模化应用，并且深入融入人们生活及工业生产制造的各个环节。特别是，互联网及移动互联网不再仅仅是人与人之间的联络通道，在物与物、物与人之间也逐步建立了持续的互动关系。部署在远端的、无处不在的传感器能够极大地扩展我们的感知能力，由这些传感器产生的数据正通过互联网改变着我们的生活。

分散部署的亿万级的各类传感器持续产生海量的数据，这些感知数据的采集、处理、传输以及存储管理、挖掘分析都让我们面临着一系列新的挑战。随着大数据技术的发展与应用，这些数据的处理与利用正得到越来越多的重视与关注。但是，不同于互联网中大数据的非结构化及价值密度低等特性，物联网应用中的数据更加倾向于结构化、及时的处理需求以及价值密度与数据量成正比等特性，需要有序的管理与处理利用。

本书作者在近几年研究开发成果及应用实践的基础上，通过系统地分析物联网中这类感知数据的特点及相应的事务处理特性，针对互联网场景下的物联网大数据提出了一套三层的物联网大数据处理的技术体系。并在此基础上，就相关的技术发展进行了深入的探讨与分析。进一步，介绍了作者所在团队研发的相关产品，以及这些产品在智能交通、智能电厂、教育、安全监控等领域的应用案例。希望这些内容能对物联网应用的开发者及两化融合、工业 4.0 环境下的大数据处理分析提供重要的指导与参考价值。

全书分为 3 篇：缘起与发展趋势篇、技术解析篇、产品研发篇。缘起与发展趋势篇包括第 1 章物联网与产业发展以及第 2 章大数据处理技术的

发展,主要针对物联网产业及大数据处理相关技术进行简要介绍和发展趋势的分析与探讨。技术解析篇包括第3~8章。第3章物联网大数据技术体系主要针对物联网大数据处理的挑战提出相应的技术体系;第4章感知数据特性与模型,进一步提出物联网感知数据库系统,并在第5章针对感知数据库系统的设计、关键技术及部署体系进行进一步的阐述;面对物联网感知数据处理的实时性需求,第6章就这类事务的实时调度、并发控制以及事务的执行模式与框架进行深入的探讨,从而为这类系统的开发实现提供有价值的参考;第7章主要针对物联网大数据在云端的存储管理进行分析;第8章主要探讨物联网大数据的计算与分析技术。产品研发篇包括第9~11章,分别介绍物联网大数据处理的三个层次中对应的产品以及这些产品的成功应用案例,为读者开展物联网大数据的应用工作提供参考。

本书主要由王桂玲、王强、赵卓峰、韩燕波编著,参与编著的还有刘晨、李寒、房俊。其中,王桂玲主要编写了第2、7、8、11章,王强主要编写了第1、3~6、9、10章,赵卓峰参与编写了第11章,韩燕波教授整体组织了本书的内容及结构,并对本书关键内容进行把关。刘晨、李寒分别参与编写了11.4节、11.5节,房俊参与编写了7.2.3节。实验室硕士研究生曹波、李冬、王路辉参与了本书部分实例的验证。

本书内容、特别是本书所介绍的相关研发产品是北方工业大学计算机学院数据工程研究院、大规模流数据集成与分析技术北京市重点实验室及中科启信公司全体人员集体努力的结晶。本书得到了国家自然科学基金(No. 61672042)、北京市自然科学基金(No. 4172018、No. 4162021)的资助。

在本书编写过程中,得到了戴国忠研究员、王晖博士的大力指导与帮助,还得到了其他许多师友的帮助和鼓励,在这里我们无法一一列举,谨向他们表示真挚的感谢。

电子工业出版社的董亚峰老师、米俊萍老师为书稿的面世给予了大力的帮助。在此,对二位老师表示衷心的感谢!

物联网大数据的应用尚处于初级阶段,由于作者水平所限,本书缺点及不足之处在所难免,欢迎广大读者批评指正。

作 者

2017年1月

目 录



第 1 篇 缘起与发展趋势篇

第 1 章 物联网与产业发展	3
1.1 物联网产业的发展	3
1.1.1 传感器与智能硬件	4
1.1.2 物联网服务平台	5
1.1.3 工业 4.0 与 CPS	5
1.2 物联网与大数据	7
1.3 物联网产业的机遇与挑战	9
1.3.1 物联网产业面临的挑战	9
1.3.2 物联网操作系统与数据库	10
1.3.3 物联网大数据处理与应用	11
第 2 章 大数据处理技术的发展	12
2.1 大数据存储和管理技术	12
2.1.1 面向大数据的文件系统	13
2.1.2 面向大数据的数据库系统	15
2.2 大数据计算技术	19
2.2.1 批处理计算模式	19
2.2.2 交互式查询计算模式	20
2.2.3 流处理计算模式	21
2.2.4 大数据实时处理的架构: Lambda 架构	23
2.3 大数据分析技术	24

2.3.1	传统结构化数据分析	26
2.3.2	文本数据分析	26
2.3.3	多媒体数据分析	27
2.3.4	社交网络数据分析	27
2.3.5	物联网传感数据分析	28
2.3.6	大数据分析技术的发展趋势	28

第2篇 技术解析篇

第3章	物联网大数据技术体系	31
3.1	物联网中的大数据挑战	31
3.1.1	互联网大数据的特征	31
3.1.2	物联网大数据的特征	34
3.2	技术体系	37
3.2.1	感知数据采集与传输	38
3.2.2	感知数据管理与实时计算	41
3.2.3	物联网平台与大数据中心	42
第4章	感知数据特性与模型	44
4.1	感知数据的特性分析	44
4.1.1	常用的感知数据类型	44
4.1.2	感知数据的主要特征	46
4.2	感知数据的表示与组织	49
4.2.1	物联网数据模型	49
4.2.2	时态对象模型	51
4.3	感知数据库的定位	52
4.3.1	感知数据库的定位	52
4.3.2	感知数据库的特征	53
4.4	感知数据库与传统数据库	53
4.4.1	感知数据库与关系数据库	53
4.4.2	感知数据库与实时数据库系统	54
4.4.3	感知数据库与工厂数据库系统	55
4.4.4	感知数据库与流数据处理系统	55

第 5 章 感知数据库管理系统	57
5.1 感知数据库的总体设计	57
5.1.1 总体设计的主要原则	57
5.1.2 感知数据库的设计框架	58
5.2 感知数据库的分布部署体系	62
5.2.1 系统的集群部署模式	62
5.2.2 多层级的系统部署体系	64
5.2.3 服务分布的部署体系	66
5.3 感知数据库中的关键技术	67
5.3.1 智能设备及传感器接口技术	67
5.3.2 流数据实时在线处理技术	68
5.3.3 事件驱动的高效处理机制	69
5.3.4 感知数据的压缩存储技术	75
第 6 章 实时事务调度处理技术	79
6.1 常见事务特性分析	79
6.1.1 感知事务	80
6.1.2 触发事务	80
6.1.3 用户事务	81
6.2 事务调度与并发控制	81
6.2.1 事务的调度方法	81
6.2.2 并发控制策略	82
6.3 服务器与操作系统	83
6.3.1 服务器体系结构与发展	83
6.3.2 操作系统的多任务机制	87
6.4 事务的执行框架与模式	90
6.4.1 通用系统模型与调度方法	91
6.4.2 事务处理框架的设计模式	91
6.5 系统框架的分析与性能优化	94
第 7 章 物联网大数据存储与管理	97
7.1 云文件系统的关键技术	99
7.1.1 HDFS 的目标和基本假设条件	99
7.1.2 HDFS 体系架构	100

7.1.3	性能保障	102
7.2	NoSQL 数据库关键技术	106
7.2.1	NoSQL 数据库概述	106
7.2.2	基于 NoSQL 数据库的物联网大数据存储与管理	118
第 8 章	物联网大数据计算与分析	123
8.1	物联网大数据批处理计算	123
8.1.1	MapReduce 的设计思想	124
8.1.2	MapReduce 的工作机制	126
8.1.3	MapReduce 在物联网大数据中的应用	128
8.2	物联网大数据交互式查询	130
8.2.1	原生 SQL on HBase	131
8.2.2	SQL on Hadoop	132
8.2.3	基于 HBase 的交互式查询	133
8.3	物联网大数据流式计算	134
8.3.1	流式计算的需求特点	134
8.3.2	流数据基本概念	135
8.3.3	流数据查询操作	140
8.3.4	流数据定制化服务	142
8.3.5	评测基准	145
8.3.6	Spark Streaming 及其在物联网大数据中的应用	146
8.4	物联网大数据分析	150
8.4.1	物联网大数据 OLAP 多维分析	151
8.4.2	物联网大数据深层次分析	157

第 3 篇 产品研发篇

第 9 章	物联网网关 CubeOne	175
9.1	工业物联网网关	175
9.1.1	CubeOne 产品概述	175
9.1.2	CubeOne 功能特点	176
9.1.3	CubeOne 的应用领域	178
9.2	无线传感器网络网关	178

9.2.1 无线传感器网络概述	178
9.2.2 ZigBee-WiFi 网关	180
9.2.3 ZigBee 网络应用案例	182
第 10 章 ChinDB 感知数据库系统	185
10.1 ChinDB 系统概述	185
10.2 ChinDB 组成与功能特点	186
10.3 ChinDB 数据组织管理	188
10.3.1 标签点及其属性	188
10.3.2 标签点的组织方式	189
10.3.3 关系数据管理	190
10.3.4 历史数据管理	190
10.4 ECA 规则与实时计算	191
10.5 ChinDB 的 HA 方案	192
10.5.1 HA 概述及模式分类	192
10.5.2 ChinDB HA 的部署模式	193
10.6 物联网应用平台	195
10.6.1 物联网平台概述	195
10.6.2 平台主要特点	196
10.6.3 应用领域与应用案例	198
第 11 章 DeCloud 物联网大数据云平台	202
11.1 DeCloud 组成	202
11.1.1 软件概述	202
11.1.2 通信服务	204
11.1.3 计算服务	206
11.1.4 存储服务	207
11.1.5 数据发布/订阅服务	208
11.2 DeCloud 在智能交通领域的应用	209
11.3 DeCloud 在教育物联网云服务平台中的应用	215
11.4 DeCloud 在电厂设备故障预警的应用	218
11.5 DeCloud 在电梯安全监控中的应用	222
11.6 DeCloud 在高精度位置服务中的应用	225
总结与展望	230
参考文献	232

第1篇 缘起与发展趋势篇

第 1 章

物联网与产业发展



经过近几年的发展，物联网从一个概念逐渐演变为蓬勃发展的新兴产业。特别是随着物联网技术在工矿企业、环境监测、智慧城市、智能家居、智能交通及智能装备等领域的广泛应用，产业发展的同时也面临着诸多的技术挑战。本章总结物联网产业的发展现状，进一步分析面临的挑战和机遇。

1.1 物联网产业的发展

当前，以移动互联网、物联网、云计算、大数据等为代表的新一代信息通信技术发展迅猛，正在全球范围内掀起新一轮科技革命和产业变革。全球物联网应用呈现加速发展态势，物联网所带动的新型信息化与传统领域走向深度融合，物联网对行业和市场所带来的冲击和影响已经广受关注，M2M（机器与机器通信）、智能汽车、智能电网、智能家居、智慧医疗等是近两年全球发展较快的重点应用领域。物联网与移动互联网多层融合协同发展，进一步带动了产业的迅速发展。

美国政府提出了制造业复兴战略，正逐步将物联网的发展和重塑美国制造优势计划结合起来以期重新占领制造业制高点。德国联邦政府在《高技术战略 2020 行动计划》中明确提出了工业 4.0 理念，推动了全球工业物联网与大数据发展的高潮。其他许多国家，如韩国政府也预见到以物联网为代表的信息技术产业与传

统产业融合发展的广阔前景，持续推动融合创新。

2015年5月，国务院印发《中国制造2025》，这是我国实施制造强国战略第一个十年的行动纲领。文件中明确了九大战略任务和重点——提高国家制造业创新能力、推进信息化与工业化深度融合、强化工业基础能力、加强质量品牌建设、全面推行绿色制造、大力推动重点领域突破发展、深入推进制造业结构调整、积极发展服务型制造和生产性服务业、提高制造业国际化发展水平。其中，新一代信息技术产业、高档数控机床和机器人、航空航天装备、海洋工程装备及高技术船舶、先进轨道交通装备、节能与新能源汽车、电力装备、农机装备、新材料、生物医药及高性能医疗器械十个产业被明确为重点发展领域。

1.1.1 传感器与智能硬件

美国IT咨询机构Gartner报告显示，由于物联网市场的迅猛发展，预计到2020年，物联网设备安装基数将突破260亿，为此，在未来几年内，物联网的发展速度将大大超过智能手机、平板电脑和PC的速度。

此外，市场研究机构BI Intelligence发布的一份预测报告显示，到2019年，由智能手机、PC、平板电脑、联网汽车和可穿戴设备组成的物联网市场规模将比现在增长一倍以上；物联网将为全球经济增加1.7万亿美元的价值，其中包括硬件、软件、安装成本、管理服务和因为实现物联网效率而增加的经济价值。

智能手表、智能手环等可穿戴产品随着Apple Watch的推出，更多地引起人们的关注。而在谷歌2014年收购Nest公司之后，围绕家庭空间的智能设备开始受到更多创业者和投资者的关注。其中，家庭安防产品国内外出现得比较多，而空气净化器产品的异军突起则是国内智能硬件发展的特例。智能路由器作为智能家居及家庭安防的操控平台中枢不断推陈出新，智能插座、智能灯泡也都开始慢慢地常态化，价格透明化。以Ghost为代表的四轴飞行器、汽车“智能穿戴”产品，以及无人驾驶汽车、智能机器人连绵不断地进入人们的视野。

赋予苹果手机及可穿戴产品越来越强大功能的，不仅是越来越强大的芯片，更重要的是手机上越来越多、越来越精良的传感器，包括触摸屏、陀螺仪、加速度传感器、红外传感器、光照传感器、指纹传感器、摄像头等。

传感器通常由敏感元件和转换元件组成，能将检测感受到的信息按一定规律转换成为电信号输出，以满足信息的传输、处理、存储、显示、记录和控制等要求。可以说，是传感器让物体有了触觉、味觉和嗅觉等感官。不仅仅是手机及可

穿戴设备,在汽车、家用电器、机器人和工业自动化领域,各式各样的传感器正成为无处不在的神经元。近年来,物联网产业的发展导致传感器需求大增,所以,传感器制造业发展很快。传感器产品品种繁多,具有高附加值,本身价格并不高。传感器的研发属于技术密集型,具有多样性、边缘性、综合性和技艺性,需要多学科、多种高新技术配合。

1.1.2 物联网服务平台

市场研究机构 BI Intelligence 预测,到 2019 年,来自物联网硬件销售的收入将只有 500 亿美元,这仅占物联网市场总收入的 8%;软件厂商和基础设施公司将分走最多份额的收入。这说明,比硬件产品更重要的是内容和服务。

作为物联网的关键支撑点,M2M 市场非常活跃,正成为全球电信运营商的重要业务增长点。中国移动在物联网领域进行了“云—管—端”的全面布局,2014 年正式发布物联网开放平台——OneNet,并且将随后推出一系列自主研发和定制开发的通信模组、穿戴式硬件和行业终端。

2014 年 8 月苹果 WWDC 大会上发布的 HomeKit 平台,主要为智能硬件开发者提供 iOS 上的数据、控制接口,实现利用苹果设备作为智能家居的控制中枢。三星在 2014 年 CES (国际消费电子展) 上首次展示了 Smart Home 智能家居平台,并且三星即将在全球范围内发布这一平台。这就意味着可以通过下载一个 Android 应用程序,来实现与三星云服务的结合,用户就可以控制智能电器。

独立的物联网平台公司也不断涌现,通过提供端到端的物联网云平台服务,帮助智能硬件制造商聚焦自己的核心竞争力,快速打造物联网产品,并且能够让任何设备实现安全连接、大数据分析,为最终用户提供功能丰富的体验。

此外,面向行业的物联网服务平台也提供了无限潜能。其中,车联网是市场化潜力最大的应用领域之一,可以实现智能交通管理、智能动态信息服务和车辆智能化控制的一体化服务。全球车载信息服务市场非常活跃,成规模的厂商多达数百家,颇具代表性的全球化车载信息服务平台有通用的安吉星 (OnStar)、丰田的 G-book。

1.1.3 工业 4.0 与 CPS

美国早在 2006 年就提出了虚拟网络—实体物理系统或者信息物理系统

(Cyber-Physical System, CPS) 的概念, 以此将物联网和互联网与制造业的融合做出综合性的概括, 并将此项技术体系作为新一代技术革命的突破点。任何产品如汽车、飞机、船舶、电梯、机床及生产线等, 都可以存在于虚拟和实体两个世界, 在虚拟世界中将实体的状态及实体之间的关系透明化, 虚拟世界中代表实体状态和相互关系的模型和计算结果能够精确地指导实体的活动, 从而使实体的活动相互协调优化。

前美国总统奥巴马在 2010 年签署了《美国制造业促进法案》, 提出运用数字制造和人工智能等未来科技重构美国的制造业优势。2012 年 2 月, 美国国家科技委员会发布了《先进制造业国家战略计划》报告, 将促进先进制造业发展提高到了国家战略层面。同年 3 月, 前美国总统奥巴马提出创建“国家制造业创新网络 (NNMI)”, 以帮助消除本土研发活动和制造技术创新发展之间的割裂, 重振美国制造业竞争力。2012 年 11 月, 美国通用电气公司 (简称 GE) 发布《工业互联网——打破智慧与机器的边界》报告, 开始向全世界推广工业互联网模式。

工业 4.0 的概念由德国在 2011 年的汉诺威工业博览会上第一次提出。德国于 2013 年正式发布了“工业 4.0 实施建议”, 拉开了全球范围内推进第四次工业革命的序幕, 如图 1-1 所示。“工业 4.0”的核心就是信息物联网和服务互联网与制造业的融合创新。报告指出, “工业 4.0”会将智能技术和网络投入到工业应用中, 从而进一步巩固德国作为生产地及制造设备供应国和 IT 业务解决方案供应国的领先地位。

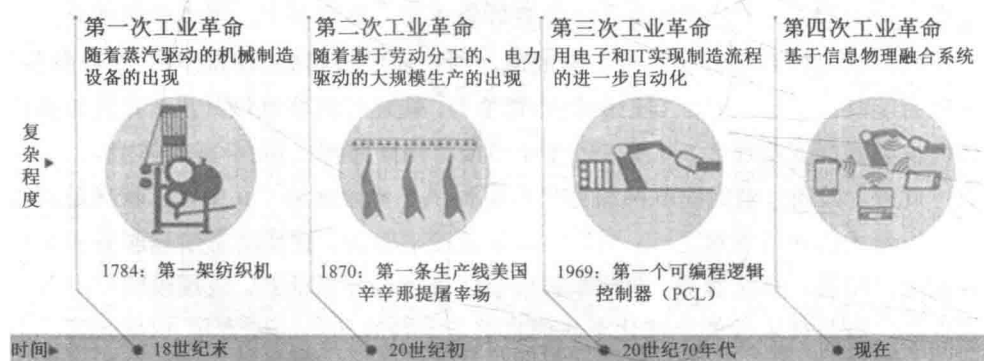


图 1-1 工业革命

美国与德国面对制造业未来虽然提出的概念不同, 但“工业 4.0”与 CPS 本质上是异曲同工的, 其战略核心是制造智能化。“工业 4.0”与 CPS 的目标在于通过物联网、信息通信技术与大数据分析, 把不同设备通过数据交互连接到一

起,让工厂内部,甚至工厂之间都能成为一个整体,在自动化之上,形成制造的智能化。以智能制造为主导的第四次工业革命,主要是指通过物联网和信息物理系统技术,将制造业向智能化转型,实现集中式控制向分散式增强型控制的基本模式转变,最终建立一个高度灵活的个性化和数字化的产品与服务生产模式。其中,物联网、互联网服务及大数据是“工业4.0”的基础。

德国工业4.0背后的战略意图,一方面是对抗美国互联网产业从“信息”领域加速进入“物理”业务领域的影响,或者说是德国希望阻止信息技术不断融入制造业之后带来的支配地位。由于CPU、操作系统、软件及云计算技术等信息技术的核心几乎都由美国企业掌控,特别是,Google开始进军机器人领域、研发自动驾驶汽车;Amazon进入手机终端业务,开始实施无人驾驶飞机配送商品。这一趋势对制造业的破坏性影响只是时间问题,工业4.0希望用“信息物理系统”升级“智能工厂”中的“生产设备”,使生产设备因信息物理系统而获得智能,使工厂成为一个实现自律分散型系统的“智能工厂”,而互联网或者云平台不过是制造业中的一个环节,不会成为整个生产制造过程的中枢。

另一方面是压制中国制造业。中国制造业的快速发展使得德、日等国的制造业感受到了巨大的威胁,工业4.0战略的推进也是德国保持国际竞争力的重大举措。中国实施《中国制造2025》,加速从制造大国向制造强国的转变。而《中国制造2025》主要是侧重于产业与政策,工业4.0主要侧重于技术与模式。不过,共同点还是智能制造。

伴随德国工业4.0的发展,生产制造领域的工业机器人将不断升级为智能机器人。作为制造业大国和机器人大国,日本也适时推出了机器人国家战略规划。2015年1月,日本政府公布了《机器人新战略》,该战略首先列举了欧美与中国的技术赶超,互联网企业向传统机器人产业的涉足,而给机器人产业环境带来了剧变。这些变化,将使机器人开始应用大数据实现自律化,使机器人之间实现网络化,物联网时代也将随之真正到来。

1.2 物联网与大数据

近年来,随着互联网的飞速发展,特别是随着电子商务、社交网络、移动互联网及多种传感器的广泛应用,以数量庞大、种类众多、时效性强为特征的非结