



数联网 ——大数据如何互联

Internet of Data
— How big data interconnect

段云峰 鄂海红 编著





国之重器出版工程
网络强国建设

网络基础与关键技术研究丛书



数联网 ——大数据如何互联

Internet of Data

—How big data interconnect

段云峰 鄂海红 编著



人民邮电出版社
北京

图书在版编目 (C I P) 数据

数联网：大数据如何互联 / 段云峰, 鄂海红编著

— 北京：人民邮电出版社, 2018.8

(国之重器出版工程·网络基础与关键技术研究丛书)

ISBN 978-7-115-48809-1

I. ①数… II. ①段… ②鄂… III. ①互联网络—应用—数据传输—研究 IV. ①TN919.1

中国版本图书馆CIP数据核字(2018)第143076号

内 容 提 要

大数据发展日新月异, 开放的数据将服务整个社会, “万数互联”的快速发展推动着数联网 (Internet of Data) 的发展。本书介绍了数联网的基本概念, 数联网的数据表示、传输、交换、跨域互操作等标准化问题, 数联网架构和管理技术, 数联网商业模式变革, 数联网应用创新等。

本书介绍的技术结合了丰富的场景故事, 内容通俗易懂, 实用性强, 适合大数据行业从业者阅读。

◆ 编 著 段云峰 鄂海红

责任编辑 李 静

责任印制 杨林杰

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

固安县铭成印刷有限公司印刷

◆ 开本: 710×1000 1/16

印张: 23.75

2018年8月第1版

字数: 440千字

2018年8月河北第1次印刷

定价: 129.00元

读者服务热线: (010) 81055488 印装质量热线: (010) 81055316

反盗版热线: (010) 81055315

《国之重器出版工程》

编辑委员会

编辑委员会主任：苗圩

编辑委员会副主任：刘利华 辛国斌

编辑委员会委员：

| | | | | |
|-----|-----|-----|-----|-----|
| 冯长辉 | 梁志峰 | 高东升 | 姜子琨 | 许科敏 |
| 陈 因 | 郑立新 | 马向晖 | 高云虎 | 金 鑫 |
| 李 巍 | 李 东 | 高延敏 | 何 琼 | 刁石京 |
| 谢少锋 | 闻 库 | 韩 夏 | 赵志国 | 谢远生 |
| 赵永红 | 韩占武 | 刘 多 | 尹丽波 | 赵 波 |
| 卢 山 | 徐惠彬 | 赵长禄 | 周 玉 | 姚 郁 |
| 张 炜 | 聂 宏 | 付梦印 | 季仲华 | |



专家委员会委员（按姓氏笔画排列）：

- 于 全 中国工程院院士
- 王少萍 “长江学者奖励计划”特聘教授
- 王建民 清华大学软件学院院长
- 王哲荣 中国工程院院士
- 王 越 中国科学院院士、中国工程院院士
- 尤肖虎 “长江学者奖励计划”特聘教授
- 邓宗全 中国工程院院士
- 甘晓华 中国工程院院士
- 叶培建 中国科学院院士
- 朱英富 中国工程院院士
- 朵英贤 中国工程院院士
- 邬贺铨 中国工程院院士
- 刘大响 中国工程院院士
- 刘怡昕 中国工程院院士
- 刘韵洁 中国工程院院士
- 孙逢春 中国工程院院士
- 苏彦庆 “长江学者奖励计划”特聘教授



- 苏哲子 中国工程院院士
- 李伯虎 中国工程院院士
- 李应红 中国科学院院士
- 李新亚 国家制造强国建设战略咨询委员会委员、
中国机械工业联合会副会长
- 杨德森 中国工程院院士
- 张宏科 北京交通大学下一代互联网互联设备国家
工程实验室主任
- 陆建勋 中国工程院院士
- 陆燕荪 国家制造强国建设战略咨询委员会委员、原
机械工业部副部长
- 陈一坚 中国工程院院士
- 陈懋章 中国工程院院士
- 金东寒 中国工程院院士
- 周立伟 中国工程院院士
- 郑纬民 中国计算机学会原理事长
- 郑建华 中国科学院院士



- 屈贤明 国家制造强国建设战略咨询委员会委员、工业和信息化部智能制造专家咨询委员会副主任
- 项昌乐 “长江学者奖励计划”特聘教授，中国科协书记处书记，北京理工大学党委副书记、副校长
- 柳百成 中国工程院院士
- 闻雪友 中国工程院院士
- 徐德民 中国工程院院士
- 唐长红 中国工程院院士
- 黄卫东 “长江学者奖励计划”特聘教授
- 黄先祥 中国工程院院士
- 黄 维 中国科学院院士、西北工业大学常务副校长
- 董景辰 工业和信息化部智能制造专家咨询委员会委员
- 焦宗夏 “长江学者奖励计划”特聘教授



前言

我们已经从 IT 时代快速迈进了 DT 时代。在 DT 时代，万数互联的社会如何发展，本书将对其进行详解。

目前，大数据在国内甚是火热，许多企业在完成内部数据的沉淀和积累过程后，逐步进入数据的开放与共享阶段。互联网的哲学，即“开放与共享”。封闭的数据仅服务于企业本身，而开放的数据将服务于整个社会，并能带来更显著的效益。数联网（Internet of Data）的概念就是在这样一个时代产生的。数联网是以互联网为承载网络，通过标准的格式、接口、协议，进行数据的交换和互联，以更加丰富的数据维度，实现数据联通之后的各种业务应用。数联网将打通企业彼此间的壁垒，完成数据的互联互通，盘活企业的数据资产，这种升级是一种质变！

推动和实施国家大数据战略最重要的方面是以数据整合和共享为途径，推动实现技术融合、业务融合和数据融合，打通信息壁垒，实现跨层级、跨地域、跨系统、跨部门、跨企业的协同管理和服务。大数据的国家战略布局和实施，无处不体现着“万数互联”的建设目标。

数联网将“站”在互联网的肩膀上，推进互联网向更深层次发展，带动社会发生本质变革。

主要内容

第 1 章探讨了数联网在演进过程中所引发的各种变革和思考，希望给阅读者以启示。该章阐述了数联网产生的历史背景、大数据到数联网演进的必然性、数联网



的概念和架构、数联网的关键内容、数联网与互联网的关系、数联网发展的意义及其多行业中的应用等内容。回顾过去，互联网的作用是解决人类信息交流问题，而数联网则面向数据的连接，解决数据的互通问题；数联网在数据连接的基础上，提供各种标准化的计算能力和更加丰富的应用，驱动企业和社会的本质变革。

数据的内容千差万别，数联网需要首先解决数据的联通问题。本书的第2章探讨和分析了数联网构建的基础——数联网的标准化问题，包括数据表示的标准化、数据传输的标准化、数据交换的标准化、数据应用接口的标准化、跨域数据互操作的标准化等内容。这种数据的联通，就像构筑高速公路的过程：给出基础建设的要求，给出高速公路上的运行标准要求，促成不同车辆在高速公路上规范、高速运行，大大提升车辆的通行率。

本书的第3章深入浅出地探讨了数联网的技术架构问题。首先，我们比照互联网的七层架构设计数联网的层次架构；其次，我们详细讲解如何解决数联网互联中数据的标准表述、数据标准化寻址、数据的浏览、数据的应用、数据的安全等相关问题。

数联网带来了新的管理问题。本书的第4章围绕数联网的生态圈探讨了数据管理架构的构建和演变。除了单点数据管理架构中的元数据管理、数据质量管理、数据标准、数据字典、数据安全、数据采集、传输、加工、处理、使用等内容，数联网管理还涉及从单点的数据管理架构延伸到多点链接的数联网生态体系的数据管理问题，如数联网中的域名、开源工具、数据编辑、参与主体等内容。

每一种创新都源于应用的需求。本书的第5章介绍了数联网的业务应用。数联网将催生新的搜索引擎、新的电商、新的社交网络、新的计算网络，让大数据的应用场景随着数据维度的丰富而产生更多的想象空间。这将对现有的互联网格局产生颠覆性的影响。

数联网的发展会面临各种各样的问题。本书的第6章探讨数联网面临的相关问题及其解决方案。在数联网时代，升级的不仅是互联网技术，更是商业模式的突破。如果说物联网是个万亿元级的市场，那么数联网也可以达到万亿元级的规模，甚至更加庞大，因为物联网也是数联网的一种数据输入源。因此，在如此庞大的规模下，如何借鉴已有的方法提出适用于数联网的解决方案是本章的核心。

数联网将促进电信行业的“凤凰涅槃”。本书的第7章描述了数联网对电信行业的影响。数联网解决了信息的互联问题，它为电信运营商进行电信大数据领域的合作拓展创造了条件。如何将电信的网络运维经验移植到数联网，实现数据价值的最大化；如何将数联网与边缘计算相结合；如何利用数联网完成电信行业的华丽转型。数联网为电信行业提供了“涅槃”的可能，运营商的数据积累也为数联网的发



展奠定了基础，电信行业与数联网是相辅相成、互惠互利的。

数据的互联将开创新时代，为世界插上“智慧”的翅膀，更为中国带来了战略机遇。本书的第8章深入地分析了中国如何从互联网大国到数联网强国。今天的数联网如同几十年前刚刚萌芽的互联网、十余年前的物联网，即将朝气蓬勃地发展，未来的数联网将会以更加成熟的姿态展现在世界面前。

本书的读者对象

- 正在经历大数据变革或准备大数据变革的制造业、服务业等企业人士；
- 大数据行业的工作人士；
- 大数据相关专业的本科生和研究生；
- 具有一定的大数据知识，并想进一步提高的人士；
- 进行大数据规划的企业、政府人员；
- 对大数据感兴趣，想了解大数据基础知识的人士。

致谢

感谢北京邮电大学计算机学院的鄂海红副教授与我共同撰写本书。感谢中国移动的各位领导和同事，其中徐海勇、陶涛、黄岩、尚晶、徐萌等为本书的编写提供了参考和建议。

感谢北京邮电大学计算机学院的白杨（第3、4章）、张田宇（第2、6章），中国移动通信集团有限公司的叶青（第7章）在书籍编写过程中所做的贡献。

感谢北京邮电大学 PCN&CAD 中心的曹师久、江裕锋、刘庆、秦景坤、张城城、张崇宇、赵文骏在数联网相关研究的过程中所做的贡献。感谢东方国信公司的黄晓宇、史富伟、刘焱等同志的辛苦工作。

本书在编写的过程中，得到了公司领导和业界朋友的各种帮助，也希望以此书回报各位的关心和照顾。

最后，感谢所有在大数据和数联网实践路上共同努力的同事朋友们，让我们共同迎接数联网的来临！



目 录

| | |
|---------------------------|-----|
| 第 1 章 概述 | 001 |
| 1.1 互联网之后是什么 | 002 |
| 1.1.1 互联网如何升级 | 002 |
| 1.1.2 大数据面临的挑战 | 004 |
| 1.1.3 封闭是“背道而驰” | 007 |
| 1.1.4 大数据的“智慧”如何联网 | 010 |
| 1.2 从大数据到数联网 | 013 |
| 1.2.1 历史溯源 | 013 |
| 1.2.2 万数互联 | 020 |
| 1.2.3 如何让应用丰富起来 | 022 |
| 1.2.4 让大数据流动起来，带来效益 | 024 |
| 1.2.5 共享经济下的数据共享 | 025 |
| 1.3 数联网的关键内容 | 026 |
| 1.3.1 数据如何互联 | 026 |
| 1.3.2 构建数联网要解决的关键问题 | 030 |
| 1.3.3 如何保护数据的隐私 | 033 |
| 1.3.4 从“免费”到“盈利” | 034 |
| 1.4 数联网的概念和架构 | 035 |
| 1.4.1 数据的标准化 | 036 |
| 1.4.2 数据访问的标准化 | 037 |



| | | |
|-------|----------------------------------|-----|
| 1.4.3 | 数据解释的标准化 | 038 |
| 1.4.4 | 数据质量的标准化 | 039 |
| 1.4.5 | 数联网的架构分享 | 040 |
| 1.4.6 | 计算的标准化 | 043 |
| 1.5 | 数联网与互联网的关系 | 044 |
| 1.5.1 | 数联网解决“数据+计算”的问题 | 044 |
| 1.5.2 | 互联网解决底层连接的问题 | 045 |
| 1.5.3 | 数联网解决数据互通的问题 | 045 |
| 1.5.4 | 数联网是互联网的升华 | 045 |
| 1.5.5 | 数联网具有更大的市场价值 | 046 |
| 1.5.6 | 是互联网上的应用？还是另一种互联网（互联网 2.0） | 046 |
| 1.5.7 | 数联网与语义网的关系 | 047 |
| 1.6 | 数联网的意义 | 048 |
| 1.6.1 | 数据的互联，将开创新时代 | 048 |
| 1.6.2 | 数联网将为世界插上“智慧”的翅膀 | 049 |
| 1.6.3 | 数联网将带动新的产业机遇 | 050 |
| 1.6.4 | 开辟新的商业模式 | 054 |
| 1.6.5 | 大数据的“倍增器” | 055 |
| 1.6.6 | 数联网的产业机会 | 056 |
| 1.7 | 数联网在各个行业中的应用 | 060 |
| 1.7.1 | 金融行业 | 060 |
| 1.7.2 | 政府决策 | 061 |
| 1.7.3 | 电信行业 | 063 |
| 1.7.4 | 汽车行业 | 064 |
| 1.7.5 | 工业 4.0 | 066 |
| 1.7.6 | 食品行业 | 066 |
| 1.7.7 | 影视行业 | 068 |
| 1.7.8 | 其他行业 | 069 |
| 第 2 章 | 数联网基础 | 071 |
| 2.1 | 定义及特点 | 072 |
| 2.1.1 | 范围 | 072 |
| 2.1.2 | 定义 | 073 |
| 2.1.3 | 与物联网间的关系，从信息链接到数据链接及应用 | 076 |



| | | |
|-------|-----------------------------|-----|
| 2.2 | 数联网的标准化层次 | 077 |
| 2.2.1 | 数联网叠加在互联网上 | 077 |
| 2.2.2 | 如何从 TCP/IP 转向数据的标准层级 | 078 |
| 2.2.3 | 数联网的标准化层次架构 | 079 |
| 2.2.4 | 数联网的 RFID——设备数据探针 /OS | 080 |
| 2.3 | 数据表示的标准化 | 082 |
| 2.3.1 | XML 的探索 | 082 |
| 2.3.2 | CWM 表述的元模型 | 083 |
| 2.3.3 | 数据的标准模型 | 086 |
| 2.3.4 | 简单的标准表格 | 087 |
| 2.3.5 | 还有哪些方法 | 087 |
| 2.4 | 数据传输的标准化 | 088 |
| 2.4.1 | 架在 TCP/IP 之上 | 088 |
| 2.4.2 | 如何借鉴 HTTP | 089 |
| 2.4.3 | 已有数据传输的标准 | 089 |
| 2.4.4 | 降低传输要求, 在应用层解读数据 | 090 |
| 2.5 | 数据交换的标准化 | 090 |
| 2.5.1 | 缘起 XML | 090 |
| 2.5.2 | 数据总线的新生 | 091 |
| 2.5.3 | 数据交换的理解 | 092 |
| 2.5.4 | 数据接口如何标准化 | 092 |
| 2.6 | 数据应用的标准化开发接口 | 094 |
| 2.6.1 | SQL 的启示 | 094 |
| 2.6.2 | 能否像堆积木一样使用数据 | 094 |
| 2.6.3 | 数据开发的语言会是什么样子 | 095 |
| 2.6.4 | 数据的传输、调用模式 (同步、异步等) | 099 |
| 2.7 | 数联网访问终端的标准化 | 103 |
| 2.7.1 | 从手机开始的数联网终端 | 103 |
| 2.7.2 | 遵循网络协议, 适合展现数据及分析 | 104 |
| 2.7.3 | 适合可视化展现 | 104 |
| 2.7.4 | 满足交互需求 | 106 |
| 2.8 | 如何打开数据库 | 106 |
| 2.8.1 | 为什么要打开数据库 | 108 |
| 2.8.2 | ODBC/JDBC 接口 | 109 |



| | | |
|--------|----------------------|-----|
| 2.8.3 | 数据库结构如何外露 | 110 |
| 2.8.4 | 数据库的应用如何外露 | 110 |
| 2.8.5 | 如何解决语义层面的互通 | 111 |
| 2.9 | 如何打开 Hadoop 等 | 112 |
| 2.9.1 | 数据分析层的通信协议 | 113 |
| 2.9.2 | 单系统的数据云化到数联网中的云化 | 115 |
| 2.9.3 | 不同生态中的数据流动 | 116 |
| 2.9.4 | Hadoop 数据如何标准化、交换和开发 | 116 |
| 2.9.5 | 数联网如何流动 Hadoop 中的数据 | 117 |
| 2.9.6 | 数据移动或计算移动 | 118 |
| 2.10 | 哪些技术可以帮助我们 | 119 |
| 2.10.1 | 如何借助现有技术进行组合 | 121 |
| 2.10.2 | 如何将中间件技术应用到数联网中 | 125 |
| 2.10.3 | 大数据的操作系统 | 126 |
| 第 3 章 | 数联网的技术架构 | 127 |
| 3.1 | 数联网基础元件 | 130 |
| 3.1.1 | 从设备感知到数据感知 | 131 |
| 3.1.2 | 数据互联“嵌入”组件 | 135 |
| 3.1.3 | 大数据 OS | 136 |
| 3.2 | 数联网的七层架构 | 140 |
| 3.2.1 | 网络的层次架构如何引申到数联网 | 143 |
| 3.2.2 | 数联网的层级架构 | 144 |
| 3.2.3 | 数联网的互联 | 146 |
| 3.2.4 | 数联网的 TMF | 146 |
| 3.2.5 | 如何打破 IOE 的垄断 | 151 |
| 3.3 | 数据的标准表述——“世界语” | 154 |
| 3.3.1 | 从“一流的企业做标准”说起 | 154 |
| 3.3.2 | 数据格式如何表达 | 154 |
| 3.3.3 | 数据的内容如何表述 | 155 |
| 3.3.4 | 数据单位的表述 | 157 |
| 3.3.5 | 数据的标准表述方式 | 158 |
| 3.3.6 | 数据的标准实现方式 | 159 |
| 3.4 | 数据的寻址 | 161 |
| 3.4.1 | 数据的地址 | 161 |



| | | |
|--------|-------------------------|-----|
| 3.4.2 | 数据如何寻址 | 164 |
| 3.4.3 | 数据的云化寻址空间 | 166 |
| 3.5 | 数据的交换 | 167 |
| 3.5.1 | 商品交换才产生价值 | 167 |
| 3.5.2 | 数据交换的标准 | 172 |
| 3.5.3 | 数据交换的内容 | 172 |
| 3.5.4 | 交换的接口如何标准化、统一化 | 174 |
| 3.5.5 | 能否直接打开数据库 | 175 |
| 3.6 | 数联网的“浏览器” | 178 |
| 3.6.1 | 互联网的浏览器及原理 | 178 |
| 3.6.2 | “超文本”之后的“超数据” | 179 |
| 3.6.3 | 数联网如何浏览数据 | 179 |
| 3.6.4 | 数联网的“网站” | 180 |
| 3.7 | 数据的应用 | 184 |
| 3.7.1 | API 能否全球统一 | 184 |
| 3.7.2 | 数据应用的模块化表述 | 188 |
| 3.7.3 | 数据的中间件 | 195 |
| 3.7.4 | 数据的共享开发 | 200 |
| 3.7.5 | 数据挖掘算法如何共享 | 206 |
| 3.8 | 数据“云化”之后的管控 | 210 |
| 3.8.1 | 与云计算的关系 | 210 |
| 3.8.2 | 数据存放在哪里 | 211 |
| 3.8.3 | 数据如何确保安全 | 212 |
| 3.8.4 | 数据质量的管理方法 | 212 |
| 3.8.5 | 数据运维的方法 | 213 |
| 3.9 | 数据安全 | 214 |
| 3.9.1 | 传统网络安全管控 | 214 |
| 3.9.2 | 数据安全技术 | 215 |
| 3.9.3 | 数据安全管理机制 | 219 |
| 3.10 | 大数据操作系统——设备的嵌入式结构 | 220 |
| 3.10.1 | 网络标准之后的终端标准 | 221 |
| 3.10.2 | 操作系统新生：从面向设备到面向数据 | 221 |
| 3.10.3 | 解决设备层面的数据传输 | 221 |
| 3.10.4 | 解决数据标准研究 | 222 |



| | | |
|--------------|------------------|------------|
| 3.10.5 | 解决数据汇聚和分析问题 | 222 |
| 3.11 | 计算能力的互联 | 225 |
| 3.11.1 | 数据互通之后，如何计算 | 225 |
| 3.11.2 | 云计算的新演化 | 226 |
| 3.11.3 | 数据仓库的变革 | 227 |
| 3.11.4 | 企事业 IT 的变革 | 227 |
| 第 4 章 | 数联网的管理架构 | 231 |
| 4.1 | 数据管理范围和内容 | 232 |
| 4.1.1 | 数据标准管理 | 232 |
| 4.1.2 | 数据质量管理 | 233 |
| 4.1.3 | 元数据管理 | 233 |
| 4.1.4 | 数据运维管理 | 234 |
| 4.1.5 | 数据安全的管理 | 235 |
| 4.1.6 | 数据流量监控 | 236 |
| 4.1.7 | 数据资产管理 | 236 |
| 4.2 | 单点的数据管理架构 | 237 |
| 4.2.1 | 从血缘影响图开始 | 237 |
| 4.2.2 | 数据处理的流程和管控点 | 238 |
| 4.2.3 | 数据如何运维 | 239 |
| 4.3 | 数联网的管理 | 240 |
| 4.3.1 | 域名的管理如何升级 | 240 |
| 4.3.2 | 大数据的管理如何升级到数联网管理 | 241 |
| 4.3.3 | 开源环境如何管理 | 241 |
| 4.3.4 | 需要维护的内容 | 242 |
| 4.3.5 | 需要制定的标准 | 243 |
| 4.3.6 | 从电信网络管理中学习到哪些内容 | 244 |
| 4.3.7 | 入网检测工作 | 245 |
| 4.3.8 | 数联网和电信网络的关系 | 245 |
| 4.4 | 人人参与数据管理 | 246 |
| 4.4.1 | 自己对自己的数据更了解 | 246 |
| 4.4.2 | 理清数据质量、数据权限 | 247 |
| 4.4.3 | 建立促进人人参与的机制 | 247 |
| 4.4.4 | 数据管理从单点到网络 | 248 |



| | | |
|--------------|--------------------------|------------|
| 4.4.5 | 更多的应用、更多的价值 | 248 |
| 4.5 | 数据质量的评估标准 | 249 |
| 4.5.1 | 从丢包开始的质量评估体系 | 249 |
| 4.5.2 | 评估数据质量 | 250 |
| 4.6 | 数联网的生态圈 | 251 |
| 4.6.1 | 网络设备制造 | 252 |
| 4.6.2 | 数据安全、数据治理 | 254 |
| 4.6.3 | 数据收集企事业 | 255 |
| 4.6.4 | 数据应用企事业 | 256 |
| 4.6.5 | 数据管理企事业 | 257 |
| 4.6.6 | 行业解决专家 | 258 |
| 第 5 章 | 数联网的业务应用 | 261 |
| 5.1 | 应用驱动创新 | 262 |
| 5.1.1 | 曾经的互联网“泡沫”改变了我们的生活 | 262 |
| 5.1.2 | 数联网的典型应用场景 | 264 |
| 5.2 | 数据交易市场 | 268 |
| 5.2.1 | 数据交易的合法性 | 268 |
| 5.2.2 | 数据交易的模式 | 269 |
| 5.2.3 | 数据交易盘活数据资产 | 273 |
| 5.2.4 | 数据交易的资本驱动 | 274 |
| 5.3 | 新的盈利模式 | 275 |
| 5.3.1 | 后向收费的场景 | 275 |
| 5.3.2 | 后向收费的好处 | 275 |
| 5.3.3 | 其他收费模式 | 276 |
| 5.4 | 应用的内容形式 | 276 |
| 5.4.1 | 数据的提取 | 276 |
| 5.4.2 | 报告的形成 | 277 |
| 5.4.3 | 专家诊断的提出 | 277 |
| 5.4.4 | 行百里、半九十 | 278 |
| 5.4.5 | 核心：结果驱动 | 278 |
| 5.4.6 | 数据编辑 | 279 |
| 5.4.7 | 在线数据服务 | 280 |
| 5.5 | 应用的 App 化 | 281 |
| 5.5.1 | 如何借鉴 App | 281 |
| 5.5.2 | 如何超越 App | 282 |