



架构师十年匠心力作 铸就科技精品

大数据

架构之道与项目实战

常耀斌 郑智民 周贤波 著

全程大项目实战，全栈工程师的技术蓝宝书，全面涵盖技术生态

5 大技术体系
SpringBoot
Hadoop
Spark
机器学习
设计模式

8 个通用平台
微服务的架构
设计、开发流
程、经验总
结，并提供全
部代码讲解

15 年顶级项目
的架构设计
思想传承，
核心代码实
战技巧的手
把手传授

30 项开源技术
的核心原理、
代码实战、核
心架构以及性
能调优的深入
阐述



清华大学出版社

内容目录

常耀斌 郑智民 周贤波 著

大数据

架构之道与项目实战



清华大学出版社
北京

内 容 简 介

大数据和人工智能技术发展正当时，如何快速构建一个高水平的企业级大数据平台是撰写本书的出发点。

本书从总体技术要求出发，深入分析了全栈技术的各自优势和应用场景，传授了三十多种主流技术的架构设计、技术原理和集成方法。第1章介绍企业级大数据平台服务的总体设计，突出研究经典设计模式之美、吸纳分布式技术的精髓、深耕微架构的演变内涵。第2章～第9章是项目实战环节，介绍高并发采集、灵活转发、高可扩展海量存储、高并发海量存储、高可靠海量存储、实时计算、智能分析和自定义迁移等微服务，手把手传授架构设计和核心代码，让读者掌握商用微服务产品开发全流程。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

大数据架构之道与项目实战/常耀斌，郑智民，周贤波著. —北京：清华大学出版社，2018
ISBN 978-7-302-51586-9

I. ①大… II. ①常… ②郑… ③周… III. ①数据处理-研究 IV. ①TP274

中国版本图书馆 CIP 数据核字（2018）第 257183 号

责任编辑：杜春杰

封面设计：常雪影

版式设计：楠竹文化

责任校对：马军令

责任印制：丛怀宇

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦 A 座 邮 编：100084

社 总 机：010-62770175 邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印 装 者：三河市君旺印务有限公司

经 销：全国新华书店

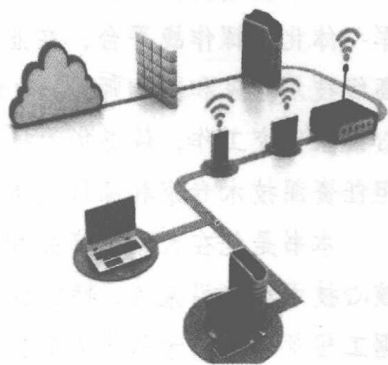
开 本：185mm×260mm 印 张：24.75 字 数：509 千字

版 次：2018 年 12 月第 1 版 印 次：2018 年 12 月第 1 次印刷

定 价：75.00 元

产品编号：081565-01

推荐序一



我国高度重视大数据在经济社会发展中的作用，十八届五中全会提出“实施国家大数据战略”，国务院印发《促进大数据发展行动纲要》，全面推进大数据发展，加快建设数据强国。大数据技术和应用正处于创新攻关期，国内市场需求正处于爆发期，大数据产业面临前所未有的发展机遇。推动大数据平台建设和智能应用发展，加快传统产业数字化、智能化，是支撑国家战略的重要举措。当前我国正在推进供给侧结构性改革和服务型政府建设，加快实施“互联网+”行动计划，为大数据产业创造了广阔的市场空间，是我国大数据产业发展的强大内生动力。

我国大数据产业具备了良好基础，目前正在内蒙古创建国家大数据综合试验区，为产业发展奠定了坚实的基础研发和商业环境，但是发展机遇和挑战并存，主要挑战有四点：一是大数据资源的开放程度和大数据质量不高，难以被快速抽取和挖掘利用；二是技术创新能力有差距，我国在分布式计算架构、大数据处理等核心技术研究上没有长时间积累，尤其是目前在大数据的开源技术上没有太多的国家影响力，我们要向谷歌等顶级互联网公司学习优秀的研发经验和创新技术；三是大数据应用的普适性存在差距，目前的大数据应用在部分垂直领域有了较深的发展，但是绝大多数领域还是空白；四是大数据人才难以满足企业发展需求，大数据基础技术研究、产品研发和业务应用等各类人才短缺，难以满足发展需要。

加强大数据技术和通用服务产品研发应以行业垂直应用为导向，推动产品和解决方案研发及产业化，创新技术服务模式，形成可以商用的、完备的技术产品体系。大数据技术，从产品体系研发上说，包括大数据采集、传输、存储、管理、处理、分析、应用、可视化和安全等关键技术；从基础技术和技术引擎上说，包括大规模异构数据融合、集群资源调度、分布式文件系统、通用计算框架技术、流计算、图计算等；从前沿技术创新上讲，包括机器学习、深度学习、认知计算、区块链和虚拟现实等。总之，打造大数据核心产品，建立完善的大数据工具型、平台型和系统型产品体系，形成成熟大数据解决方案，是大数据产业化的重点工作方向。

本书作者是我的研究生，他于2007年获得北京邮电大学计算机专业硕士学位，曾

任职于中国电科，参与并负责一体化指挥平台的核心服务研发工作，历时2年研制了全军一体化指挥作战平台，在业内取得了良好声誉。2009年加入华为技术，担任云平台高级技术专家和架构师，主导设计了国内首个企业公有云平台，负责云计算和大数据核心技术研究工作，获得华为技术“金牌员工”和“总裁奖”。现就职于中国移动研究院，担任资深技术专家和高级技术经理，专注于大平台建设和微架构等核心技术研发。

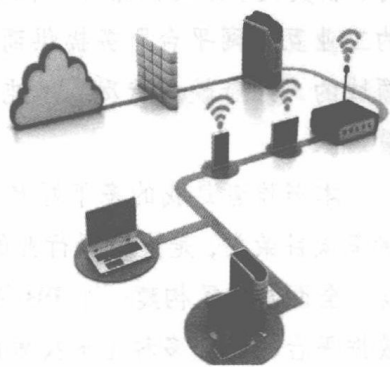
本书是他在大数据产品和研发技术上的十多年研究成果，总结了通用大数据平台的核心技术和前沿架构，特别适合大数据从业者，尤其适合高等院校的毕业生、在职大数据工程师和有志于从事大数据研发的工程师阅读，也可以作为工具类和实战类技术手册使用。本书是大数据全栈工程师的摇篮。

本书介绍并实践了大数据技术，包括8个通用微服务开发，是作者从事架构师工作多年的实战经验总结。本书精心总结了多种创新设计思想，全面阐述了8个通用服务引擎的开发过程，涵盖了高并发采集服务、灵活转发服务、高可扩展海量存储服务、高并发海量存储服务、高可靠海量存储服务、实时计算服务、智能分析服务和自定义迁移服务。每一位读者都可以借助这些微服务快速开发出个性化的平台和应用。

“在科学的道路上没有平坦的大道，只有不畏艰险沿着陡峭山路向上攀登的人，才有希望达到光辉的顶点。”希望作者和每一位读者都能求真务实，努力攀登数据科学的高峰，为新时代的科技发展贡献自己的智慧和力量！

北京邮电大学博士生导师 邓中亮教授

推荐序二

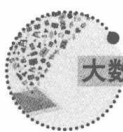


大数据概念最早由世界领先的咨询公司麦肯锡提出,之后短短几年在全球范围工业界、学术界、商业界等获得巨大的应用和推动。在美国硅谷大量的公司都投身其中,老牌的 IT 公司,如谷歌的 APP Engine、微软的 Azure 等;初创公司更是层出不穷。越来越多的公司意识到大数据蕴含的商业机会,信息社会的数据爆炸也不断促使了大数据技术向前飞跃发展,然而大数据对技术的高要求和数据本身的私密性决定了大数据不是人人都有资格参与,必须具备相应的技术储备和资质才可能投身其中。

随着全球数据的爆发式增长,大数据在中国也从政策层面备受关注。2014 年,大数据首次写入政府工作报告,大数据逐渐成为各级政府关注的热点,政府数据开放共享、数据流通与交易、利用大数据保障和改善民生等概念深入人心。2015 年 8 月 31 日,国务院印发《促进大数据发展的行动纲要》,成为中国发展大数据产业的战略性指导文件。作为我国推动大数据发展的战略性、指导性文件,充分体现了国家层面对大数据发展的顶层设计和统筹布局,为中国大数据应用、产业和技术的发展提供了行动指南。2016 年,《中华人民共和国国民经济和社会发展第十三个五年规划纲要》正式公布,“十三五”规划纲要对国家大数据战略的阐释,成为各级政府制订大数据发展规划和配套措施的重要指导,对我国大数据发展具有深远的意义。2016 年年底,工信部正式发布《大数据产业发展规划(2016—2020 年)》,该发展规划以大数据产业发展中的关键问题为出发点和落脚点,明确了“十三五”时期大数据产业发展的指导思想、发展目标、重点任务、重点工程及保障措施等内容,成为大数据产业发展的行动纲领。

大数据技术究竟包括哪些?大数据技术是一个交叉学科,涉及云计算、物联网、人工智能等技术,可以在各行各业得到相关的应用,如环保、医疗、人口、能源交通等各个领域,可以共性地概括为凡是涉及数据生产、采集、存储、加工、分析等行为的领域,包括数据资源建设、大数据硬软件产品的开发等工程实现的都包含在大数据技术的范畴。

围绕国家大数据战略实施要求,一些国内知名的互联网领头企业,如华为、中兴、阿里、百度、腾讯等软硬件企业陆续推出大数据相关平台和产品。以蚂蚁金服、滴滴出



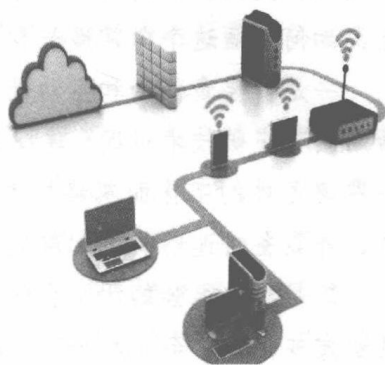
行、新美大、菜鸟网络等为代表的新兴独角兽企业，以及以树根互联、徐工信息为代表的工业互联网平台服务提供商也纷纷布局大数据领域。但是各公司都实际面临大数据领域的人才奇缺的状况，单纯依靠高校培养大数据人才远远不够，需要更多的社会力量加入。

本书作者是我的多年好友，他带领团队完成了国内多个国家级云平台建设，获得了诸多项目荣誉，是大数据行业的知名专家。他从大数据平台的架构设计核心技术要求出发，全面阐述了构建一个 PB 级规模数据运营能力的云平台的具体要求，零距离传授大数据平台的三十多种主流技术的架构设计、技术原理、集成方法、性能调优，全面涵盖 Spring 生态、Hadoop 离线计算、Spark 实时计算、机器学习和设计模式五大生态领域，注重理论和实际项目相结合，代码量丰富、详尽。

在当前大数据领域书籍偏重理论分析、缺少实际项目论述的背景下，这本书显得尤其实用，不仅可以成为广大希望投身大数据领域的从业者一本很好的引路书籍和工具，而且也可以作为高等院校相关专业学生学习大数据的参考资料。

国资委网络安全专家 北京科技委专家 张 伟
大唐集团科研院网络安全实验室主任

前言



大数据在近几年逐渐步入理性发展期。未来的大数据发展会渗透到各个垂直行业领域，尤其在企业数据资源化、传统行业智能化、分析引擎产品化等方向上迎来前所未有的发展机遇。与此同时，大数据的中高端人才缺乏严重阻碍大数据产业发展，尤其是掌握技术、精通管理、拥有项目经验的高级工程师成为企业炙手可热的人才。在北上广深杭等一线城市，有一年以上工作经验的大数据工程师的月薪过万，有几年工作经验的数据分析师的年薪在 30 万~50 万元之间，而更顶尖的大数据技术人才则是年薪超百万，成为各大互联网和一流企业争夺的对象。2018 年 1 月，208 所高校、高职院校获批开设“大数据技术与应用”专业。同时，教育部设立了“数据科学与大数据技术”本科专业，35 所大学已获批开设，还有 200 多所高校正在申报。纵观企业的大数据人才需求和高校的课程设立布局，应试者的实操能力和项目经验积累越来越受到注重。在专业技能上，重点需求体现在大数据基础平台建设和应用开发上，涵盖了高并发采集、灵活转发、海量存储、实时计算、智能分析和高效迁移等专业技能。在个人能力要求上，重点需求体现在责任心、团队合作能力和解决问题能力等方面。本书重点讲授大数据全栈专业技能，但接下来，我先谈谈大数据工程师的个人能力要求和职业发展规划。

我从事计算机领域研发 16 年来，主导负责了十多个国家项目，经历了单机服务到分布式服务的项目研发模式，也实践了从百万用户到上亿用户的商用产品。但是，在新的移动互联网和大数据时代，对工程师和技术研发人员的专业技能和个人能力提出了新的要求，单兵作战模式无法成就一个商业产品，需要设计驱动和团队协同作战。团队合作和协同作战一直是我倡导的软件产品商用化的管理模式，我的很多学生在 BAT 等知名互联网公司担任技术主管，从某种意义上来说，得益于我很早就对团队赋予了设计引领产品的创新思想。但是，在我的职业发展中也曾多次面临发展瓶颈，我是如何面对挑战和压力呢？接下来，我想分享一下在软件研发职场上的晋升技巧和有效工作方法，让大家少走弯路，多获捷径。大数据工程师的职业发展路线大致分几个关键阶段：一是上升为项目经理阶段；二是历练为技术经理阶段；三是发展为资深架构师阶段；四是成长为首席技术官（CTO）阶段。

第一个阶段是上升为项目经理，先争取在项目中担任技术骨干，并逐步主动承担和

肩负更多更具挑战的研发任务。项目经理职位在一流的互联网公司至少需要奋斗3年以上,如何缩短这个非常漫长和艰苦奋斗的过程,建议从以下几个方面做起。

一是需要责任心和主动性,不仅要按时完成项目经理交付的开发任务,最好还经常帮助同事突破技术难题。建议是:一定要为成为技术专家而不懈努力,千万别奔波在做一些事务性的工作而忽略技术本身,软件工程师的核心竞争力就是拥有全面的核心技术,并具备快速解决技术难题的能力。

二是要有很强的团队合作能力,善于发现别人优点并学会适当表扬,善于总结自己的研发成果并学会主动分享,善于表达自己并学会归纳总结。团队合作能力也是需要不断提升的,多听取别人的忠告而改变自己,多帮助别人解决问题而感受快乐,多用心学习核心技术而不搬弄是非。

三是要学会成就团队,就是能创造一个环境,让每个人都能在其中发挥出更多的能力,也就是一种领导力。我在华为工作期间,主动承担了一些技术维护工作,经常为大家管理服务器并配置环境,很快被同事们赋予“大管家”称号,自己在享受称号的同时也得到了领导的认可,虽然技术维护工作本身不计算在绩效中,但是其对团队的运营能力提升是举足轻重的。学会成就团队是树立你在团队中威望的很重要品质。

四是要提升汇报和总结能力。不管是民营企业还是国有企业,汇报能力对职场人都是非常重要的,究其原因汇报能体现一个人的综合能力,需要有严谨的逻辑思维和优秀的写作能力,让领导在短时间内掌握一个项目的开发现状、存在问题、解决方案和创新工作等。更好地规划工作、布局工作、超预期完成工作,并在适当时候提出有建设性的宝贵意见是至关重要的。汇报工作的核心是分析和解决方案,领导都是团队中最忙和承担压力最大的人,比起发现问题来说,他更关注的是问题分析和最优解决方案。工程师学会多思考问题并有针对性地提出优秀解决方案,对团队和个人发展都是至关重要的。比如我们的项目因缺乏设计而导致开发周期太长,比如我们的项目因不能定期和客户沟通需求而导致偏离实际需求,比如我们不能按期交付项目成果而导致领导不满意。

第二个阶段是历练为技术经理。这个阶段对于一般的项目经理而言就是一个项目接着一个项目交付,上升空间遥不可及,如果要突破晋升空间也是有工作方法和拓展思路,建议从以下几个方面做起。

一是把控好项目的里程碑并学会提升管理水平。项目要有合理规划,从项目工作计划到项目任务分解、从技术选型到技术验证成功、从总体设计规划到架构设计细化、从架构设计分解到概要设计说明、从概要设计到详细设计落实、从详细设计规划到核心代码编写等,都是需要不同阶段的技术评审和质量审查,都是需要分时段交付研发成果,都是需要管理和技术能够协同推进。

二是要加强团队建设,更关注人才的能力和培养。带团队就是带人心,在公司规则之内多考虑员工的合理想法,切不能顾此失彼地加压。从团队建设力度就可以看出公司



的发展动向，如果以人为本重视长期发展，常常会考虑激励和培养员工，这恰恰也是小公司的努力方向，希望多重视程序员的意见并采纳实施，其实这样做之后最大受益者也是老板。

三是要多输出具备影响力的项目成果，如商用产品、项目奖项、核心专利和高水平学术论文。任何公司都是需要产品布局和发展规划的，尤其是短时间内要占领行业制高点，最好是有核心竞争力的产品或者专利来支撑公司在行业的领航地位。这个核心竞争力就是来源于项目而高于项目本身的拳头产品。我们在研发中多积累优秀设计思想、多总结提炼核心算法、多琢磨技术难题的创新解决方案、多讨论问题碰撞思想火花、多研究论文学习前辈的前瞻思想。

四是多读书，多学习优秀管理思想，领会分层管理的领导艺术和对结果负责的管理体制，更不能越权管理。技术经理往往需要掌握分级分层的管理思想。如果我们的日常工作都聚焦在具体事务上，如果不关注产品运营而拘泥于任务细节上，如果没有远大的理想和成就一番事业的抱负，如果没有带领团队打造核心产品的目标，那么在行动和执行力上就会出现小格局、小思维，最终因为延误战机而失去创造奇迹的机会，没有成功的团队就不会有成功的个人，没有成功的个人谈何脱颖而出的成功技术领导人？华为公司成功的原因之一就是层层管理者都要保证按期交付而不越权管理，高层领导负责战略和市场，中层领导负责战术落地实施并跟踪任务，基层人员负责细节实施和按期交付，一个完备的权责明确的分层管理机制一定会推动公司高效的运营。越权管理不仅会导致基础管理员失去权力而懈怠，更会导致不能细化管理而延误进度，大目标都是小里程碑积累完成的，不积跬步何以至千里。

第三个阶段是发展为资深架构师。这个阶段需要在知名企业的一线产品上历练十年以上，架构师是一个既需要掌控整体，又需要洞悉局部瓶颈的技术领袖。架构师在整个产品研发的生命周期中都起着至关重要的作用，随着开发进程的推进，其职责或关注点不断地加深。在需求分析阶段，软件架构师主要负责梳理非功能性系统需求，如软件的高可维护性、高性能、高复用性、高可靠性、有效性和可测试性等，另外，架构师还要经常分析客户不断变化的需求，确认开发团队所提出的设计；在总体设计阶段，架构师的关注点主要在开发团队的技术能力和开发模式；在软件概要和详细设计阶段，架构师负责对整个软件体系结构、关键构件、接口和开发策略的设计；在代码编写阶段，架构师则成为详细设计者和代码编写者的老师，并且要经常性地组织一些技术研讨会、技术培训班等来提升团队的技术能力；在软件测试交付阶段，架构师跟踪关注性能需求，同时开始为下一版本的产品是否应该增加新的功能模块进行决策。从架构师的工作职责上说：一是必须具有丰富的软件设计与研发经验，并验证所进行的设计是如何映射到实现中去；二是要具有领导能力与团队协作能力，架构师必须是一个团队最核心的技术领导人，能在关键时刻对技术的选择做出及时、有效的决定；三是要有不断积累新技术和新

架构的技术能力，架构师需要掌握的知识是多维度和多方面的，如精通各种标准的通信协议、网络服务、面对对象数据库、关系数据库或者 NoSQL 数据库、数据处理和分析等知识，另外，架构师应与时俱进地学习新软件设计和开发思想，并不断探索更有效的新方法。开发语言、设计模式和开发平台不断地升级，架构师需要吸收这些新技术、新知识，并将它们用于软件产品和项目开发工作中。总的来说，架构师是一个技术高端职位，技术经理如何得到这样的机会，如何利用所掌握的技能进行应用的合理构架，如何不断地抽象和归纳自己的构架模式，如何深入行业成为一流公司的架构师，确实需要不断地磨炼。

第四个阶段是成长为公司 CTO。这个阶段需要有敢为人先的胆识、阅历丰富的见识、与时俱进的学识。作为 CTO，要想突破自己的领导位置，要敢于大胆提出创新思想和超前理念来带领团队脱颖而出，同一个起跑线上更需要与众不同的决心和勇气，好的机会永远是留给有胆识的技术领导人，切勿在关键时刻瞻前顾后或者犹豫不决，否则很难成就一番大业，最有说服力的例子就是 BAT 的创始人，他们用超人的胆识抓住了移动互联网高速发展契机，各自打造自己擅长的垂直领域，通过核心技术让本地化和移动设备完美结合，改变了新时代下的人的消费和社交模式。要想突破自己的技术职级，要提升自己对行业信息和外界发展的见识，不能守旧在自己的技术领域。很多技术经理习惯停留在技术舒适区，不愿意进入挑战区，而且很少参加国际或者国内的主流峰会，很难提出高瞻远瞩的创新性的解决方案。作为技术领导人，要想超越同行成为佼佼者，要静下心来沉淀和历练，只有拼出来的美丽，没有等出来的辉煌。

有了如上的职业奋斗目标，我们要脚踏实地地开始走入大数据研发工作中了，下面是本书章节安排。

- 第 1 章：大数据平台服务总体设计，从总体技术要求阐述微服务的核心需求和设计方法。
- 第 2 章：深入阐释“高并发采集微服务”的架构设计和技术实现。
- 第 3 章：着重论述“灵活转发微服务”的架构设计和技术实现。
- 第 4 章：详细介绍“高可扩展海量存储微服务”的架构设计和技术实现。
- 第 5 章：重点讲述“高并发海量存储微服务”的架构设计和技术实现。
- 第 6 章：重点说明“高可靠海量存储微服务”的架构设计和技术实现。
- 第 7 章：详细论述“实时计算微服务”的架构设计和技术实现。
- 第 8 章：主要介绍“智能分析微服务”的架构设计和技术实现。
- 第 9 章：深入论述“自定义迁移微服务”的架构设计和技术实现。

本书的创作成果是我多年的架构经验积累和核心技术提炼，从开始撰写到完成初稿大概花费了一年多时间，我钻研了上百本大数据技术专著，希望把这些前瞻技术运用于项目实战，让大数据工程师们快速学到真本领，把知识转化为科技生产力。在此，我首



先感谢我的团队和同事们，一起走过了科研之路，一起孵化了商用产品，一起收获了专业技能。

感谢郑智民、周贤波、于路、马超、曹权红、郭义华、王威等技术专家多年对我工作的积极支持和热心帮助。尤其感谢曹权红，他年轻有为，勤奋好学，专注于大数据技术研究，在项目研发上已经能独当一面，是团队中的技术达人和奋斗者。马超也是团队中的佼佼者，在技术上积累了丰富的研发经验。

感谢我的导师邓中亮教授，传授给我很多探索科学的方法，是我从事科研工作的领路人，作为一名著名科学家，他获得了诸多国家科技荣誉，但一直奋斗在科研一线，非常值得我敬重。他能在百忙中抽出时间给我作序，非常感谢。

感谢我的家人，我的妻子和女儿，是她们对我的科研工作的默默支持，才让我能全力以赴地投入到技术研发上。

感谢我的父母，他们不仅给了我生命，更多的是培养我锲而不舍地去攀登科学高峰的品质。

感谢对我的书进行推荐的朋友们，是你们对我的支持才有了这本书创作的源泉。

感谢我的学生们，你们的青春都是用来奋斗的，相信大数据产业发展的未来属于你们。

特别说明：本书代码都是基于前瞻技术 Spring Cloud 分布式架构设计，读者如有研究需要，可以联系作者。

常耀斌

2018年10月于北京



第 1 章 企业级大数据平台服务的总体设计

1.1	平台架构设计的总体技术要求	2
1.2	微服务引擎的可扩展性设计	6
1.3	微服务引擎的优秀解决方案	8
1.3.1	高并发采集微服务	11
1.3.2	灵活转发微服务	12
1.3.3	高可扩展海量存储服务	13
1.3.4	高并发海量存储服务	13
1.3.5	高可靠海量存储服务	15
1.3.6	实时计算服务	16
1.3.7	基于机器学习的智能分析服务	17
1.3.8	自定义迁移服务	17
1.4	设计小结	17

第 2 章 大数据高并发采集微服务引擎

2.1	核心需求分析和优秀解决方案	20
2.2	服务引擎的技术架构设计	20
2.2.1	Maven 与 Eclipse 集成配置	22
2.2.2	Mina2.0 框架以及业务设计	24



2.2.3	设备协议规范制定及数据包设计	26
2.2.4	按照设备和数据类型进行业务树构建	30
2.2.5	按照设备的数据包状态进行解析	32
2.2.6	按照通用方式进行高并发入库	34
2.3	核心技术讲解及模块化设计	35
2.3.1	Spring Maven Web 服务构建	35
2.3.2	Spring Boot 微服务构建	42
2.3.3	数据包定义和实现	51
2.3.4	业务树构建和实现	62
2.3.5	数据包状态进行解析实现	79
2.3.6	按照通用方式进行高并发入库实现	90
2.3.7	客户端模拟器工具类进行高并发测试	126
2.4	项目小结	139

第 3 章 大数据灵活转发微服务引擎

3.1	核心需求分析和优秀解决方案	142
3.2	服务引擎的技术架构设计	142
3.3	核心技术讲解及模块化实现	145
3.3.1	Spring MVC Web 服务构建	145
3.3.2	Spring Boot 微服务构建	152
3.3.3	灵活配置和通用工具类构建	156
3.3.4	创建发送数据主题，注册观察者对象	159
3.3.5	启动多线程进行数据发送	165
3.3.6	采用 Post 策略模式进行数据发送	168
3.3.7	采用 ActiveMQ 策略模式进行数据发送	169
3.4	项目小结	173

第 4 章 大数据高可扩展海量存储微服务引擎

4.1	核心需求分析和优秀解决方案	176
-----	---------------------	-----

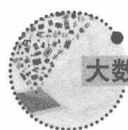
4.2	服务引擎的技术架构设计	177
4.3	核心技术讲解及模块化实现	179
4.3.1	Spring MVC 的工作原理及执行流程	179
4.3.2	Spring MVC Web 服务构建	180
4.3.3	Spring Boot Web 微服务构建	187
4.3.4	统一对外数据接收接口及通用类	191
4.3.5	MySQL 对智能终端运动数据的分状态和分策略处理	203
4.3.6	MySQL 对智能终端运动数据的分职责处理	210
4.3.7	MySQL 对智能终端运动数据的统一入库处理	214
4.4	项目小结	228

第 5 章 大数据高并发海量存储微服务引擎

5.1	核心需求分析和优秀解决方案	230
5.2	服务引擎的技术架构设计	230
5.3	核心技术讲解及模块化实现	231
5.3.1	Spring MVC 和 Spring Boot 集成 MongoDB	232
5.3.2	MongoTemplate 核心类实现 Dao 层接口	234
5.3.3	基于 MongoDB 处理智能终端运动数据	236
5.3.4	基于 MongoDB 管道技术处理体检数据	243
5.3.5	基于 AngularJS 架构可视化体检数据	264
5.4	项目小结	270

第 6 章 大数据高可靠海量存储微服务引擎

6.1	核心需求分析和优秀解决方案	272
6.2	服务引擎的技术架构设计	272
6.3	核心技术讲解及模块化实现	273
6.3.1	Hadoop 完全分布式集群构建	274
6.3.2	Spring MVC 和 Spring Boot 集成 Hbase	279
6.3.3	HbaseTemplate 核心类实现 Dao 层接口	280



6.3.4	Hbase 集群的智能终端运动数据 Controller 接口	286
6.3.5	Hbase 集群的智能终端运动数据 Service 接口	288
6.3.6	Hbase 集群的智能终端运动数据 Dao 接口	292
6.4	项目小结	300

第 7 章 大数据实时计算微服务引擎

7.1	核心需求分析和优秀解决方案	302
7.2	服务引擎的技术架构设计	302
7.3	核心技术讲解及模块化实现	303
7.3.1	分布式采集服务 Flume 部署及数据采集	303
7.3.2	分布式消息服务 Kafka 部署及数据发送	305
7.3.3	创建 HBase 数据库和 Spark 环境	307
7.3.4	分布式实时处理引擎 Spark Streaming 原理及数据处理	308
7.3.5	构建 BD_RTPServer_DP 工程实现数据处理	309
7.3.6	构建 BD_RTPServer_Boot 服务实现可视化	318
7.4	项目小结	327

第 8 章 大数据智能分析微服务引擎

8.1	核心需求分析和优秀解决方案	330
8.2	服务引擎的技术架构设计	330
8.3	核心机器学习算法讲解和应用	331
8.3.1	逻辑回归的原理分析	331
8.3.2	支持向量机原理分析	333
8.3.3	决策树原理分析	334
8.3.4	聚类算法原理分析	336
8.3.5	关联规则算法原理分析	336
8.3.6	协同过滤原理分析	336
8.4	Spark 架构原理与数据预测	337
8.4.1	YARN 运行架构工作原理	339

8.4.2	Spark Mlib 核心技术	341
8.4.3	Spring Maven 工程构建	342
8.4.4	决策树预测体检费用	345
8.4.5	逻辑回归预测体检费用	347
8.4.6	随机森林预测体检费用	349
8.4.7	支持向量机预测疾病概率	350
8.4.8	协同过滤推荐药品	352
8.5	项目小结	353

第 9 章 大数据自定义迁移微服务引擎

9.1	核心需求分析和优秀解决方案	356
9.2	服务引擎的技术架构设计	356
9.3	核心技术讲解及模块化实现	357
9.3.1	Hadoop 生态的核心组件	357
9.3.2	HBase 工作原理	358
9.3.3	Sqoop 工作原理	360
9.3.4	MapReduce 工作原理	360
9.3.5	Sqoop 抽取历史数据到 HDFS	361
9.3.6	构建工程 BD_CustomTransfer_Maven	364
9.3.7	智能终端运动数据从 MySQL 数据迁移到 Hive	374
9.4	项目小结	377