a mindstyle business 与思想有关

为数据而生
大数据创新实践

周 涛◎著

BIG DATA
Innovation

2015年度十大科技创新人物 周涛 首部个人专著

北京联合出版公司
Beijing United Publishing Co.,Ltd.

为数据而生

大数据创新实践

周涛◎著

BIG DATA
Innovation

北京联合出版公司
Beijing United Publishing Co.,Ltd.

图书在版编目（CIP）数据

为数据而生：大数据创新实践 /周涛著. —北京：北京联合出版公司，2016.5

ISBN 978-7-5502-7583-6

Ⅰ. ①为… Ⅱ. ①周… Ⅲ. ①数据处理 Ⅳ. ①TP274

中国版本图书馆CIP数据核字（2016）第073014号

上架指导：经济趋势 / 大数据

为数据而生：大数据创新实践

作　　者：周　涛

选题策划：果库文化

责任编辑：徐　樟　徐秀琴

封面设计：红杉林文化

版式设计：果库文化　李晓红

北京联合出版公司出版

（北京市西城区德外大街 83 号楼 9 层　100088）

北京鹏润伟业印刷有限公司印刷　新华书店经销

字数 163 千字　720 毫米 ×965 毫米　1/16　14.5 印张　1 插页

2016 年 5 月第 1 版　2016 年 5 月第 1 次印刷

ISBN 978-7-5502-7583-6

定价：52.90 元

献给我的父母，
他们的爱让我长大。
献给我的爱人，
她的爱让我永远不需要长大。

Big Data
Innovation
自序

在麻瓜和魔法师之间作出选择

我在中科大读本科的时候，上过一门关于“符号计算”的课程。当时授课的老师跟我们说，她以前曾经花很多年的工夫学习和研究过“怎么样在以穿孔纸带为输入方式的计算机上高效实现一些数值计算”。当时她的研究水平和成果在国内应该是领先的，本以为就靠此成就大业了，但是很快，这个世界上突然就再也找不到博物馆以外的穿孔纸带了——我们现在都用键盘和鼠标了。

纸带机的故事让我想起了一个有些悲伤的段子，我且用第一人称复述一遍。我有一个表哥，因为盗窃被抓。表哥负责藏赃销赃，团伙其他人不知道脏物在哪里，他也死活不承认自己知道，结果被重判了 10 年监禁。坐牢之后老婆也跑了，亲戚朋友也散了，只有我还时不时去看望一下，带些东西。直到快出狱的时候，表哥才跟我说，等出狱了要带我一起发大财。我当时特别激动，经常

在梦中被大富大贵的场景惊醒，也觉得自己真的是好心有好报。等表哥出狱的时候，我隆重地给他接风洗尘。表哥也迫不及待，当晚就买了两把大铁锹，拉着我去郊外一个林子里挖宝。“是金条？是银元？”我激动不已，表哥却笑而不答。我们大半夜挖出了两个大铁箱，然后用铁锹把生锈的锁头劈开，哇，满满两铁箱的传呼机。

讲这两个例子，是想说我们这个时代变化太快——**这个时代的特征就是有很多新时代层出不穷**。而咱们中国人，最最悲哀的事情，就是经常以为自己是时代的精英，最终却成了时代的弃儿。N 年以前最让人艳羡的一群人，不是大学生，而是国有企业的工人。他们或许没有想到有一天自己的“金饭碗”会被打破，贫病下岗。现在又有一大群人，削尖脑袋想挤进公务员或者事业单位人员的队伍，好一辈子守着公务员编制或者事业编制。对，就是这群扑火的人，会在未来编制改革的时候看清楚自己飞蛾的本体。

什么样的人才能在下一个时代生存和发展

那么，问题来了，什么样的人才能在下一个时代生存和发展呢？是那些拥有公务员编制或者事业编制的人吗？**在下一个时代，自动化、定量化和个性化会成为主要的特征**。恒河沙数的智能终端将会遍布这个世界——从农场到工业制造装置，从智能家居到人体内外。这些智能终端采集和产生的数据，经由数据挖掘和机器学习的手段加工分析，不仅能够提高传统农业、工业的效率，还能够为每一个人提供包括教育、零售、娱乐、金融和医疗等方面完全个性化的服务。驱动这个时代来临的关键力量是数据与数据化的思维。

拥有大数据的理念，能够掌握数据和运用数据的人，就是下一个时代的魔法师，反之，你就成了麻瓜！不管你今天从事的是什么行业，金融、医疗、教育甚至只是一个一线的产业工人或者服务人员，你所在的行业将来都很可能被颠覆，你现在的职业将来都可能变成一种自动化的服务。面对奇幻而又危险的未来世界，今天你就需要在麻瓜和魔法师之间做出选择！在一个麻瓜占绝大多数的世界里面，做一个麻瓜也没有什么不好的，然而很可能，未来的世界是一个魔法世界，你还满足于做一个麻瓜吗？

用数据说话，做最棒的魔法师

最棒的魔法师，是既深谙大数据的理念，又掌握着大数据的核心技术。但是，对于绝大部分人来说，后者是有困难的。我想特别强调的是，**即便你不能掌握一项特定的数据技术，了解大数据的理念，培养大数据的思维模式，也是非常重要的**——不管你从事什么工作，这种大数据的思维模式都是有帮助的。事实上，我一直觉得类似于统计学（包括概率论、数理统计、统计物理等）和机器学习的理念，对于我们理解这个世界都是有帮助的，应该有一些生动的科普书，把这些重要的理念用通俗的语言告诉大家。

数据化思维的核心是什么？就是定量化，或者说“用数据说话”。主观能动性当然是我们人类的重要能力，特别是行业专家的思路和判断往往非常重要，效果甚至好于机器学习的结果。但是，一切的评估都要定量化。举个例子来说，要证明一个营销行为 B 比营销行为 A 更好，必须要无偏地把用户划分成两个群，一个接受 A 一个接受 B，然后通过对比来验证两者的效果。政府做决策的时候，例如改变医保的规则，也需要充分的数据支撑，提前能够量化这个改变带来的

效果，并且时时监督政策实施后的结果。学会用数据来说明“哪个更好哪个更坏”，是数据化思维的第一步。

作出让世界尊重的原始创新

当魔法师的另一个好处，就是我们可以进入魔法世界——这是一个浪漫的战争世界，我们必须变得更强，才能打倒伏地魔！

在我读大学的时候，我们的案头枕边，放着的是茨威格的《异端的权利》，是索尔仁尼琴的《古拉格群岛》，我们追忆和供奉几千年来为了人类进步付出甚至牺牲的科学家、哲学家、文学家、政治家，等等，我们能够非常清楚地说出哪些人是世界的脊梁。我们在字里行间追寻中国最苦难最黑暗的时代，羡慕在那个时代战斗的英雄，我们急切地希望这个时代能够让我们为民族的复兴战斗——尽管可能不是用刀枪！

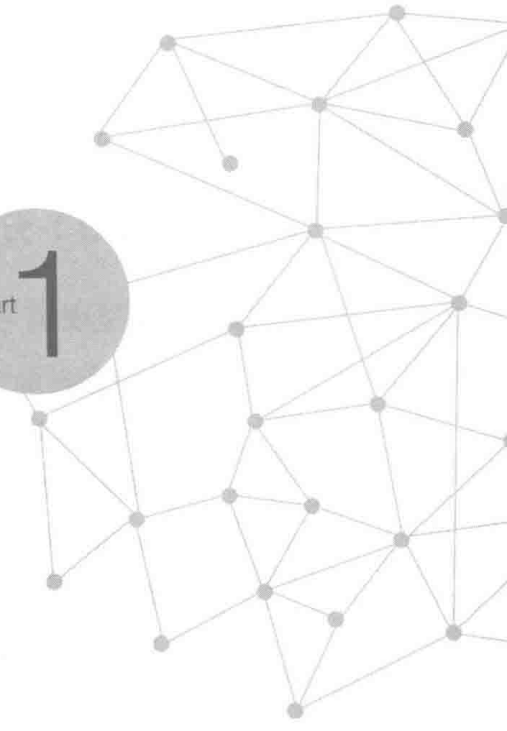
我不知道我们这一代，是不是中国流淌着战斗血液的最后一代大学生。我们现在面对的是不一样的战场，不是刺刀机枪，而是要做让世界尊重的原始创新。我在这本书里面描写了很多在大数据领域努力拼搏希望有所创新的中国人，尽管他们中的绝大部分距离成功还非常远，但我希望他们的故事以及这些故事背后的理念、技术和精神，能够唤起更多的创新者。

有两个问题，我希望每一个读者都问问自己。**第一，在你的一生中，有没有可能作出类似于 SpaceX 和 AlphaGo 这样让世界尊重的原始创新。**人生特别美好的一件事情，就是通过努力，把一件看起来不可能的事情做成！这个问题可以换一个问法，就是如果有 10 个最聪明厉害的人，愿意 3~5 年竭尽全力

为你工作，你会和他们一起做一件什么事情？**第二，你所做的事情，能够为我们的国家乃至整个世界，产生什么样的重大贡献**。建一个色情网站、开发一款暴力游戏，也能挣大钱，而且很快。致力于优化教育资源或医疗资源的配置，可能非常苦非常慢，挣钱也不如暴力游戏，但是可能改变甚至拯救一大群人。如果让我选择，我会选择后者。事实上，你所贡献的要比你所得到的更能体现你的价值！

有些了解我创业历史的人，掰着手指数我的企业和资产，几千万、几亿、几十亿……然后看着我千年不变的穿着，就认为我是一个艰苦朴素不懂得享乐的人，甚至笑话我是榆木脑袋。其实恰恰相反，我是一个非常了解生活品质，而且非常资深的吃货，也从来不觉得高级的享受是一种耻辱。我有很多非常喜欢吃的东西，而且往往都价格不菲：巴西松子、车厘子、山竹、哈根达斯朗姆酒味的冰淇淋……有的时候，我在超市里面或者路上看到这些东西，非常想吃非常想买，但是我都会问自己，我最近几天做了什么贡献，有什么成果，是否配得上去享受这些东西。绝大多数时候，我都忍住了。

序终于写完了，我去买山竹了，啦啦里啦啦。



Big Data Innovation

目录

大数据 1.0：分析

Part 2

Part 3

大数据 2.0：外化

大数据 3.0：集成

Part 4

你不是一个人在读书！
扫码进入湛庐“趋势与科技”读者群，
与小伙伴“同读共进”！

Big Data

忽如一夜春风来，千“数”万“数”梨花开。大数据这个概念突然之间席卷全球，势不可当！“荒林春雨足，新笋迸龙雏”。很多研究人员似乎感受到了春意的召唤，抖抖身子，“呼哧呼哧”就变成了大数据的专家。他们发表大数据的文章、撰写大数据的著作、提交大数据的报告、召开大数据的会议、申请大数据的项目，在一个本来纯粹而美好的概念身上喷涂了一夜暴富的泡沫。这种没有准备也没有判断的一拥而上，会使我们迷恋浅表的形式，而无法吮吸深刻的内容。新时代的宫门正缓缓开启，而我们中的大部分人，注定会一边山呼新时代万岁的口号，一边埋头冲进旁边的厕所。十年回首，先机已失，“山回路转不见君，雪上空留马行处”。本部分将教会大家分辨，何处是宫殿，何处是厕所。

Part 1

大数据时代，用数据说话

上帝创造了整数，所有其余的数都是人造的。

利奥波德·克罗内克，德国数学家和逻辑学家