



a mindstyle
business 与思想有关

BIG DATA

THE BUSINESS
REVOLUTION

车品觉◎著

大数据实践的先行者
阿里巴巴集团副总裁、数据委员会会长
车品觉 首部个人专著

决战大数据

驾驭未来商业的利器



浙江人民出版社
ZHEJIANG PEOPLE'S PUBLISHING HOUSE



BIG DATA
THE BUSINESS
REVOLUTION

决战大数据

驾驭未来商业的利器

车品觉◎著



图书在版编目 (CIP) 数据

决战大数据：驾驭未来商业的利器 / 车品觉著；—杭州：浙江人民出版社，2014.4

ISBN 978-7-213-06000-7

I. ①决… II. ①车… III. ①企业管理—数据管理 VI. ①F270.7

中国版本图书馆 CIP 数据核字 (2014) 第 035596 号

上架指导：经济趋势 / 大数据

版权所有，侵权必究

本书法律顾问 北京诚英律师事务所 吴京菁律师
北京市证信律师事务所 李云翔律师

决战大数据：驾驭未来商业的利器

作 者：车品觉 著

出版发行：浙江人民出版社（杭州体育场路347号 邮编 310006）

市场部电话：（0571）85061682 85176516

集团网址：浙江出版联合集团 <http://www.zjcb.com>

责任编辑：朱丽芳

责任校对：朱志萍

印 刷：蕺城市京瑞印刷有限公司

开 本：720 mm × 965 mm 1/16 印 张：14.75

字 数：14.6 万 插 页：2

版 次：2014 年 4 月第 1 版 印 次：2014 年 5 月第 2 次印刷

书 号：ISBN 978-7-213-06000-7

定 价：45.90 元

如发现印装质量问题，影响阅读，请与市场部联系调换。

忘掉大数据

我们正处在历史的转折点上，数据技术在快速变革。大数据成了人们竞相议论的热词，但鲜有人提及这场巨大变革中人们需要具备的能力。无数的企业及个人望“数”兴叹：“大数据与我何干？”未来是大数据的时代，未来的竞争就是数据的竞争。也许，我们早该忘掉那些华而不实的喧嚣，让大数据真正从“看”到“用”，真正“活”起来。

大数据的力量来自触类旁通的关联。我们以前总是用数据来证明或企图说服工作上的盲点，而如今的数据不再成为一加一的依据，而是具备了预测和开创新机的能力。

用数据找机会

我第一次见证大数据的魅力是在 15 年前，只是当时的我并不知道那就是“大数据”。20 世纪 90 年代初，我认识了一些以博彩为生的朋友，这些人组成了一个团队，每年通过赛马，就能盈利数亿港元。我非常惊讶，要知道很多人在赌马场上可是血本无归的，而他们却能把这种概率游戏变成稳定的盈利工具。原来，他们的秘密就是使用了一套“养数据”策略——将每一场赛马比赛的过程都录了下来。当时我觉得这个做法很奇怪：“电视上已经在

BIG DATA

The Business Revolution

前言

播放录像了啊，还另外录比赛干什么？”后来我得知，他们居然在每场比赛中都会录取赛马不同角度的录像。通过这些录像，他们分析出骑师、马匹有哪些失误动作，这些动作会带来怎样的后果，然后再把这些数据“清洗”出一个更准确的数据（Smart Data）。赛马过程中有许多意外，他们利用数据来还原——如果在没有意外发生的情况下，马匹在不同场地与不同骑师配合中的应有速度。就这样，他们可以更准确地判断出每匹马的实力和获胜的机会；就这样，通过悄无声息的数据收集，每年入账数亿港元。

令我最为震惊的是，他们竟然不看表面数据，而是从无限数据的机会中寻找核心数据。

这正体现了大数据与以前数据最大的不同。以前，我们都是有问题找数据，而大数据时代，其最核心的特质则是“用数据找机会”。我们做大数据，必须要有一个预判，那就是哪些数据是你必须要提炼出来来解决盲点的。赛马的结果其实充满了“意外”，新的数据角度帮助我们一窥真实的结果，这就是“用数据来还原真实”。

只有实效的数据才是正道

现在，大数据的概念纷繁复杂，媒体上充斥了各种关于大数据的报道，但其中不乏牵强附会、滥竽充数的言论，某些媒体甚至把简单的统计也冠上了“大数据”的头衔。

作为一个跟数据打了十几年交道的人，我深深地知道从“看”到“用”，再从“用”到“养”的运营数据，本身就是一个复杂的过程，而也许目前我们最应该做的，就是暂且忘记大数据的概念。行胜于言，只有具备实效性的数据才是正道。我希望从一个实用的角度来拨开大数据的“迷

雾”，告诉每个人大数据的具体运作应该是什么样的；而且数据量绝对不是一个最重要的问题，我们要的不是数据的量，而是有“质”的量，这正是我写作本书的重要目的。

数据，决胜未来的商业利器

在这个风云变幻的数据时代，只有让数据成为商业的利器才能决胜千里。

首先，我们需要拥有一套具有商业敏感度的数据决策框架，可以使企业“看”得更准，并能够对近期做了什么是对的、什么是错的进行判断。这样一来，快速的数据反馈可以让每个决策的误差得到适时修正。其次，让数据真正从“看”到“用”，让用数据成为构建企业生产力的重要部分。再次，让 Data Technology（DT 战略）深入到企业的每个角落，使数据从生产、收集、使用、分享到反馈变得简单易用。最后，让 DT 战略落地还要特别注意数据的稳定、准确、时效和有效实施。

在阿里巴巴，我学习到一个很重要的经验——人和事是分不开的。企业要想成为一家数据化的公司，文化的培养必不可少。“混、通、晒”及“存、管、用”两套内功是让企业的血液（数据）流动起来的关键。数据流动得宜，则神清气爽；相反，数据如果出现停滞或质量问题，企业则性命不保。

锻造数据力和思考力的合力

大数据时代的到来，让我们几千年来第一次碰上了数据化带来的机遇和滚滚红利，也让每个人的思维方式出现了重大改变。很多时候，我们欠缺的不是解决问题的方式，而是定位问题的能力；我们应该思考，

有没有数据可以改善我们的盲点；我们该如何学会用“假定数据是可获取的”来重新思考周围的一切。当这种数据化思考成为你的“直觉”时，就能够把数据的力量和思考的力量融合在一起，产生出新的无与伦比的合力了。到那时你会发现，周围的一切都将因此而焕发出新的生机。

最后，衷心感谢一年来帮助我完成这个看似不可能完成的任务、写完这本看似不可能完成的书，最终付诸实践的老师及朋友们：

感恩给予我写这本书勇气的雪谦寺拥珠扎西仁波切；

感谢一直帮助我推动这本书前期工作的林君君；

感谢用业余时间协助我完成写作及案例工作的赵军、李江、吴慧敏、欧吉良、张大红；

感谢湛庐文化专业的策划和编辑团队；

更感激家人一直容忍着我的忽略……

正如数据是一种信仰，万法唯心，每一刻的进步都来自真诚的对待！

品觉

于西子湖畔

2014年4月

• 特别说明 •

1. 本书观点乃作者十几年经验心得的分享，只代表个人立场。
2. 本书的缘起来自解决雪谦寺重建的经费及桑珠助学的资金问题，本人的所有收益将全部捐出作以上用途。

BIG DATA

The Business Revolution

目 录

前 言 /1

忘掉大数据

第一部分 从数据化运营到运营数据

01 大数据，为什么很多人只会谈，不会做 /003

大数据从来不是免费的午餐

人的断层

模型数据从何而来

更主动的管理，更多的创新

数据化思考 问题就是答案

02 大数据的本质就是还原用户的真实需求 /017

识别，让似是而非的行为数据串联起来

价值，企业价值 Vs 客户价值

场景，你知道当时所有的场景吗

还原是一个瞄准器

数据化思考 CEO 们关心哪三个数据

03 “活”的数据才是大数据 /035

“活”做数据收集，抓住相关性

“活”看数据指标，动态地使用数据

数据化思考 别再做“碰巧游戏”

04 无线数据，大数据的颠覆者 /051

无线数据正在将整个数据变成“噪音”

PC数据与无线数据的关系

多屏思维下的电子商务

数据化思考 样本的偏见

05 数据分类与数据价值，什么才是你的核心数据 /065

数据分类为什么如此重要

数据分类的4大维度

数据的5大价值

数据化思考 用傻瓜的视角去观察

06 从用数据到养数据 /087

数据应用因小而美

把数据放进“框”之中

如何用框架来做决策

养数据，重要的数据战略

数据化思考 远离“或”选择

07 数据的盲点，负面数据的力量 /103

数据盲点

小偷思维

数据盲点的价值

数据化思考 为什么数据会骗人：常态、时态与变态

第二部分 阿里巴巴的大数据秘密

08 阿里巴巴的大数据实践 /123

假定数据是稳定的

假定数据是可获取的

数据化思考 先开枪，后瞄准

09 混、通、晒，阿里巴巴数据化运营的内三板斧 /135

混，“混”出数据

通，打“通”“混”的数据

晒，“晒”出“混”和“通”的数据

数据化思考 思考，要学会关窗口

10 存、管、用，阿里巴巴运营数据的外三板斧 /153

存，数据收集的开始

管，保护好存储数据

用，从收集数据到管理数据

数据化思考 用化骨绵掌解决本质问题

11 大数据，未来商业的利器 /169

假定数据是脏的

学会慢慢淡化数据

数据的标签化管理

重要的是数据和数据之间的关系，而不是数据本身

数据的实时化与实时性分层

未来是人机的结合体

数据化思考 忽略了趋势，过去的价值一文不值

BIG DATA

The Business Revolution

第一部分

从数据化运营到运营数据

BIG DATA

The Business Revolution

数据十诫

- ① 好的问题，答案就在里面。
 - ② 在实践中提炼数据。
 - ③ 让数据（Data）变成科技（Technology），惠及更多人。
 - ④ 让数据跟着“人”走。
 - ⑤ 木有数据质量，神马数据都是浮云。
 - ⑥ 以“假定数据是可获取的”去思考问题。
 - ⑦ 大数据安全，不是监管。
 - ⑧ 利用数据拿到更有用的数据。
 - ⑨ 建立数据的数据，才有进步。
 - ⑩ 让人做人擅长做的事，让机器做机器擅长做的事。
-



01

大数据，
为什么很多人只会谈，不会做

B I G
DATA

The Business Revolution

BIG DATA 众说《决战大数据》

The Business Revolution

林奕彰 eBay 大中华区 CEO

我跟品觉相识多年。在大数据还没流行以前，他就开始关注这个趋势，研究数据应用的意义以及怎样真正读懂数据和利用数据。这本书深入浅出，把深奥的大数据讲得生动有趣，非常好读；书里面集中了很多品觉个人的真知灼见，非常值得一读。读此书是一种享受，正如我享受与品觉的友谊一样。

有人说，大数据就像国王的新衣，每个人都在国王面前说着动听的话，国王信以为真，其实竟然不知道自己在裸奔。

的确，网络上有很多人在谈大数据，但是他们只会谈，不会做，因为他们根本就没有做过，包括那些所谓的“大数据专家”，他们真的做过吗？没有。

事实上，这些人对大数据内在的问题一点儿都不了解，更别说知道大数据的水有多深了。

目前，在大数据方面，无法深入应用的原因在于，从收集到使用的大数据价值链出现了问题。从理论上来说，从收到用的螺旋式循环是一个巨大的涡轮，只有先数据化运营，然后才能运营数据。而现在的情况是，用数据的人不知道大数据从哪里来，做数据的人

不知道大数据如何使用。用的人不敢用，因为大数据的真实性；做的人不知道怎么用，因为大数据的复杂性。这一问题造成的结果就是，数据量变得越来越大，而且越来越无法有效地使用。

大数据从来不是免费的午餐

大数据从来不是免费的午餐。大数据的来源是多种渠道的，倚、随机的误差总是存在。

我先问一个数据管理上最现实的问题：“大数据如何备份？”毫不夸张地说，大数据已经这么庞大了，如果再备份一次，你的成本起码会增加一倍。

做大数据基本上都要从大量地收集数据开始，因为这些数据在未来会大有用处。但是，你是不可能无止境地收集下去的。在这里，你已经看到了一个再清晰不过的伪命题——大数据的确能够备份，但是成本会增加两三倍……然而，“以前重要的数据肯定都需要备份啊，”你自然会问，“如果不能备份，我该怎么办？”而这就是一个大数据管理中必然要遇到的问题。

我再问一个问题，如果你在数据使用方面一直得心应手，整个商业链条和数据紧密相关、相辅相成。但是，现在数据链忽然断了，或者不再有效了，你该怎么办？

你需要注意的是，这一问题说的不是你有没有使用好数据，而是说曾经你可以得到的数据现在无从获得了；或者说这些数据不能再在线上收集，只能从线下获取，成本也就相应升高了；抑或说，

这些数据存储在其他地方，你不能使用了。在诸如此类的情况下，你该怎么办？

当然，如果你对于数据的使用本来就很生疏，而且也觉得无关紧要，那么数据链消失了、断了、失效了，你也不会有太大烦恼。但假如正是你将数据使用得风生水起的时候，如果出现这样的问题，那你也就只能束手无策，眼睁睁地看着机会溜走。

大数据从来不是免费的午餐。伴随着大数据热潮的到来，关于大数据的一些新问题层出不穷——大数据会夹杂着虚假信息；大数据的数据量很大，但有用的信息不一定多，甚至还会破坏核心信息；大数据的来源是多种渠道的，偏倚、随机的误差总是存在。

但是，我们也需要客观地认识到，大数据现在面临的这些问题，其实就是把小数据中的一些问题放大了。小数据中难道就没有噪音会破坏我们的核心信息吗？当然也有，只不过当大数据把数据量放大和变多的时候，噪音的破坏性也相应增大。小数据中难道就没有渠道偏倚和随机的问題吗？当然也有，但是在大数据的背景下，问题就被更明显地放大了。

人的断层

断层才是大数据所面临的最严重的问题。现在，收集数据的人并不清楚未来使用数据的人要做什么，这是目前大数据的一大关键命门。

说了这么多，事实上还没有触碰到如今大数据面临的最大问题：人。