

人人都是产品经理
www.woshipm.com

起点学院 联合出品

Broadview®
www.broadview.com.cn

职业产品经理面试、入行和进阶学习宝典
京东大数据产品总结、大厂智慧结晶



数据产品经理 **宝典**

大数据时代如何创造卓越产品

李阳 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>



数据产品经理 **宝典**

大数据时代如何创造卓越产品

李阳 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

“数据”两个字越来越频繁地出现在大家的工作中。一方面，“用数据说话”成为每个互联网从业者必备的“生存技能”；另一方面，一个名为“数据产品经理”的职位成为各大互联网企业的“热招职位”。那么，作为数据产品经理，有了数据应该怎样“用数据说话”？又应该如何让自己具备独特的竞争优势呢？

本书内容涵盖了数据产品经理应该知道和掌握的基础知识——从每个优秀的数据产品经理都应当关注的“效率”问题出发，分别从商业知识和技术知识两个角度，针对什么是数据产品、数据产品诉求的产生和类型、数据产品的实现思路与常见技术方案等关键问题进行讲解。

本书既是学习指南，又是速查手册，适合具备不同工作背景并正在从事数据产品经理工作的人士阅读，也适合对这一领域感兴趣并希望从事数据产品经理工作的人士阅读。只要你具备求知的热情，本书将为你提供解决问题的思路、方法和工具。

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

数据产品经理宝典：大数据时代如何创造卓越产品 / 李阳著. —北京：电子工业出版社，2020.4
ISBN 978-7-121-38627-5

I. ①数… II. ①李… III. ①数据处理—产品设计 IV. ①TP274

中国版本图书馆 CIP 数据核字(2020)第 034592 号

责任编辑：林瑞和 特约编辑：田学清

印 刷：三河市鑫金马印装有限公司

装 订：三河市鑫金马印装有限公司

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：720×1000 1/16

印张：20.5

字数：378 千字

版 次：2020 年 4 月第 1 版

印 次：2020 年 4 月第 1 次印刷

定 价：69.00 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：010-51260888-819, faq@phei.com.cn。

前言

数据产品的发展已经变得不容忽视。从简单的报表，到各种可视化工具，再到人们通过各种模型来实现针对每一位用户的个性化服务，这些领域都有数据产品的影子。

数据产品的关注点是什么呢？或者说，当面临多种方案的时候，什么才能指导我们做出最终的选择呢？如果要找出本书“一以贯之”的核心词，那就是“效率”。这几乎是商业领域做数据产品的全部关注点，甚至是商业的全部关注点。产品运营关注运营的效率，业务发展关注发展的效率。此外，推广有推广的效率，投资有投资的效率，盈利有盈利的效率……

而数据产品关注的是数据应用的效率，也就是各种使用数据的场景是否足够高效、是否存在不必要的浪费、是否存在不合理的设计、是否还有提升空间等。可以说，在业务越来越依赖于数据的今天，数据应用的效率已经成为业务发展效率的重要组成部分。因此，数据产品不是“冷漠”的辅助性技术工具，不是业务“心脏里的支架”，它应当是业务的“心脏”。

肩负这样的“重任”，作为数据产品经理，除了发挥想象力，在实现效率提升的过程中也有一些具体的方向需要思考。有“科学管理之父”之称的弗雷德里克·泰勒（Frederick Taylor）就提出了“精细化、数量化、标准化”的理论来提升效率。

首先，将所有要做的事情拆分开，拆分到每一个步骤、每一个处理过程。然后，对每个步骤的执行过程进行数量化。衡量每个过程的不同实现方式的效率，衡量投入的时间、可能存在的失误等风险，最终选出每个过程最高效的完成方式。最后，将每个过程的“最佳实践”推广成为标准，让大家都能参照执行。

在本书的正文部分，笔者提供了大量的“分而治之”的、精细化的思路，既包括

对业务目标的拆解、对业务发展诉求的拆解，又包括对用户、市场和需求的拆解。这其中涉及大家经常提到的“指标拆解”“用户分群”“系统模块化”等。

针对拆解后的各个方面，笔者将统一给出衡量投入和产出的思考方式。其中，包括针对不同的产品生命周期、不同的用户生命周期、不同的产品类型和行业等因素，找到不同场景下的关注点和衡量标准，并据此制定不同的应对策略。

除了这些比较“零散”的具体分析，笔者介绍了许多经过很长时间仍然为人们所津津乐道的模型。其中既包括管理学领域的，又包括营销学领域的，当然也少不了来自互联网领域的各种分析模型。这些模型提供了标准化的分析问题的方法，并且针对各自适用的问题给出了标准化的解决方案。

大家在这一系列思维方式和模型的指导下，能够快速发现现有数据应用过程中的效率问题，并不断突破这些效率瓶颈，实现整体提升数据应用效率的目的。比如，通过整合数据存储和加工的流程，来减少重复建设和浪费。大家或许经常听到“整合”这个词，分析能力在整合、数据在整合、技术基础设施在整合，乃至整个企业或团队都在整合。

当然，大家也不能机械地套用理论，还需要结合实际情况做出权衡和匹配。比如，数据平台的长期发展与短期诉求的权衡，通用内容与定制内容的权衡；数据应用与基础数据的匹配，业务诉求与技术能力的匹配等。在本书中，笔者也会提及做出这些权衡和匹配的依据，包括成本投入、价值产出和其他影响效率的因素。

“数据产品经理”这个职位在不同的企业或团队中也有不同的定位，有的偏向数据分析，有的偏向工具平台搭建。本书尽量覆盖这两种类型的不同要求，为想要成为数据产品经理的朋友呈现一个“以数据产品驱动业务”的全景。另外也借此提醒希望成为数据产品经理的朋友，应优先了解该职位的定位及方向，再结合自己的兴趣，判断是否接受这个职位。

目标读者

本书的目标读者可以分为四类。

一、目前已经成为数据产品经理，并已经具备一些经验的人群

首先，祝贺你已经成为数据产品经理中的一员！

本书内容对于已经成为数据产品经理的人群来说，是他们熟悉的内容。作为数据产品经理，能够直接在日常工作中感受到本书中提到的各种问题和解决方案在实际中的应用。衷心希望笔者对解决方案的思考能够帮助大家在工作中取得新的突破。

二、对业务层面较为熟悉、希望成为数据产品经理的人群

那些具备分析师、运营及商科背景的人群，自身对业务层面发生的事情已经非常了解。但这几类角色对技术层面的了解又是不同的。在他们当中，最有可能接近技术层面的角色，应当是分析师。分析师更有可能接触到偏底层的数据表，以及 SQL 和数据加工过程等内容。当然，这与不同企业内的分工方式有关。有些企业也确实存在更偏向分析思路和弱化分析工具的“分析师”岗位。

相比之下，具备运营背景的人对用户和市场的感知更偏向感性和人性的层面，更了解用户的思维方式和核心诉求，而具备商科背景的人则对业务本身更了解，对业务的运作方式、未来发展方向和商业目标的达成等方面会有更充分的知识储备和掌控能力。

以上三种角色的突出优势，在数据产品的设计和搭建过程中会发挥各自不同的作用。这三种角色，对本书中的业务和分析相关内容已经非常熟悉了，因此可以重点阅读第一篇的两章内容，以及从第 5 章开始的偏向技术方向的内容。这些可能是大家在搭建数据产品的过程中遇到的主要瓶颈。

三、对技术层面较为熟悉、希望成为数据产品经理的人群

对于工程师及具备理工科背景的人群来说，那些逻辑性较强的内容已经不在话下。特别是对于一些有实战经验的工程师来说，系统中各类功能的具体实现方式已经烂熟于心。即使是专门面向数据的数据产品，除了一些数据处理中的“新情况”，其他部分已经不需要过多讨论了。

不过，由于数据产品不仅是一个“纯技术”型的产品，还有一种支撑业务、赋能业务、实现“Guide Business”的智能系统，因此对于具备理工科背景的人群来说，其中的业务分析思路和商业逻辑部分可能是相对欠缺的，大家可以重点阅读第一篇的两章内容，并重点了解第二篇中各章节的内容。

四、其他对数据感兴趣、希望成为数据产品经理的人群

如果大家对数据产品感兴趣，那么不管擅长什么，都欢迎大家更多地了解数据产品及其背后的商业和技术部分。相信这会为大家施展自己的才华提供足够的空间。

目录结构

本书共分为三篇。

第一篇，主要讲解数据产品自身的建设目的和可能遇到的问题。

- ❑ 第1章，主要介绍数据产品重点关注的效率问题，以及效率问题是如何产生的。
- ❑ 第2章，主要介绍搭建数据产品的核心目标，并通过理解业务和理解技术两个方面来辅助搭建出优秀的数据产品。

第二篇，主要讲解业务层面的问题，从业务自身的逻辑到业务与技术系统之间的关系。

- ❑ 第3章，借助“投资”思维中的投入和产出概念，分别分析了几种常见的业务的商业模式，并介绍了几种常用的分析业务的管理模型和营销组合。
- ❑ 第4章，主要讲解业务的维系与发展等方面对数据的诉求，包括用户市场方面的研究、业务及产品自身形态的研究，以及通过与数据产品深度结合实现业务综合能力升级的过程。
- ❑ 第5章，主要讲解从数据角度看到的“业务”，包括用数据抽象来自用户市场的需求，用数据抽象业务自身的逻辑，以及业务的“数据世界观”和几种针对行业的通用数据仓库模型。

第三篇，主要讲解技术层面的问题，针对技术系统对业务的支撑，介绍了几种经典的大数据技术框架。

- ❑ 第6章，主要讲解技术系统对业务的支撑方式，以及业务中的数据形态和业务运转中可能遇到的数据问题。
- ❑ 第7章，主要讲解在架构层面两种经典的产品技术架构，并讲解程序代码

层面的“做事思路”，以及从技术层面进行产品模块化的方法。

- ◇ 第 8 章，主要讲解大数据技术框架在设计时常见的关注点，这些关注点也是大数据领域常见的核心问题，同时还介绍了几种常见的大数据技术框架及其处理数据的基本逻辑。

【读者服务】



微信扫码回复：（38627）

- 获取博文视点学院 20 元付费内容抵扣券
- 获取免费增值资源
- 获取精选书单推荐
- 加入读者交流群，与更多读者互动

目录

第一篇 理解数据产品：确实有些不一样

第1章 什么是数据产品	2
1.1 数据产品的关注点	3
1.2 什么是数据应用	5
1.2.1 数据处理的角度	5
1.2.2 数据展现形式的角度	7
1.2.3 应用目的的角度	9
1.3 什么是效率问题	12
1.3.1 成本投入项	12
1.3.2 价值产出项	14
1.3.3 效率的问题	15
1.4 本章小结	17
第2章 数据产品面临的挑战	18
2.1 为什么要做——师出有名	19
2.1.1 支撑数据应用	20
2.1.2 “量入为出”的价值管理	29
2.2 做的是什么——理解业务	29
2.2.1 数据的意义	30
2.2.2 架起“量化运营”的桥梁	33
2.3 怎样做到的——理解技术	35
2.3.1 理解“究竟能做些什么”	36

2.3.2 思考“怎样做得更高效”	42
2.4 本章小结	45
第二篇 理解业务：“奇怪”的数据需求从哪来	
第3章 业务是什么	48
3.1 业务的目标是什么	50
3.1.1 能力视角	50
3.1.2 利润视角	52
3.1.3 效能视角	52
3.1.4 影响力视角	53
3.2 业务的商业模式与“投资”思维	56
3.2.1 资金投资	57
3.2.2 人力投资	63
3.2.3 时间投资	66
3.2.4 其他投资	67
3.3 常用管理模型和营销组合	68
3.3.1 常用管理模型及其关系	68
3.3.2 常用营销组合及其关系	96
3.4 本章小结	101
第4章 业务的数据诉求	103
4.1 用户市场研究	104
4.1.1 需求分析的目的	105
4.1.2 需求的分层	108
4.1.3 需求的定位	116
4.1.4 需求分析的评价与 KANO 模型	127
4.1.5 需求的传播和贯彻	129
4.2 业务及产品形态研究	130
4.2.1 评价标准——怎样才是“好”	131
4.2.2 业务转化与价值归因	144

4.2.3	流量管理与实验框架	153
4.3	综合能力升级	159
4.3.1	分析方法论及其优化	160
4.3.2	固化应用系统与赋能业务	171
4.3.3	赋能团队合作	174
4.4	工具、模型与业务、产品的“日常”	176
4.5	本章小结	179
第5章	用数据抽象业务	180
5.1	需求研究的数据抽象	181
5.1.1	需求挖掘——投放与获得新用户	182
5.1.2	需求鉴别——留存与促进用户活跃	189
5.1.3	用户生命周期与“蓄水池”模型	194
5.1.4	竞争性抽象与建模	200
5.2	业务的数据模型	204
5.2.1	用E-R图抽象实体关系	205
5.2.2	用流程图抽象业务过程	212
5.2.3	用时序图抽象处理过程	219
5.2.4	用财务思维抽象资金流	225
5.3	“数据世界观”	234
5.3.1	数据模型与现实世界的差异	234
5.3.2	用户行为的事件模型	235
5.4	数据仓库建模	242
5.4.1	面向分析的数据模型	242
5.4.2	通用数据仓库模型	244
5.5	本章小结	250

第三篇 理解技术：打开数据系统的“黑箱”

第6章	从业务诉求到技术系统	252
6.1	实现业务诉求的方式	253

6.1.1	主动反馈与被动反馈	254
6.1.2	通用内容与定制内容	256
6.1.3	离线分析与在线分析	257
6.1.4	全量与抽样数据	258
6.2	业务中的数据形态	259
6.2.1	业务理解与元数据	259
6.2.2	离线数据与数据集	260
6.2.3	实时数据与数据流	261
6.3	业务中的技术问题	263
6.3.1	数据量激增问题	264
6.3.2	如何处理“陈旧”的内容	267
6.3.3	数据安全问题	268
6.4	本章小结	272
第7章	必要的技术基础知识	274
7.1	产品的技术结构与“技术世界观”	276
7.1.1	Client/Server 结构	277
7.1.2	Browser/Server 结构	278
7.1.3	产品的“技术世界观”	279
7.2	代码理解世界的“做事思路”	280
7.2.1	面向过程	280
7.2.2	面向对象	282
7.3	系统的基本模块化	283
7.4	本章小结	284
第8章	常见大数据技术框架	286
8.1	大数据技术框架的几个关注点	287
8.1.1	多——数据量	288
8.1.2	杂——数据结构	290
8.1.3	乱——数据到达	296
8.1.4	急——时效性	299

8.2	常见大数据技术框架及基本逻辑	302
8.2.1	Apache Flume 和 Apache Kafka	303
8.2.2	Apache Hadoop	306
8.2.3	Apache Hive 和 Facebook Presto	310
8.2.4	Apache Kylin	311
8.2.5	Apache Flink 和 Apache Storm	312
8.2.6	Apache Spark	315
8.3	本章小结	316

第一篇

理解数据产品：确实有些不一样

数据产品是供给「用户」
提供数据服务的产品

数据产品的关键特征

数据产品是供给「用户」提供数据服务的产品。数据产品的关键特征包括：数据驱动、用户导向、持续迭代、数据安全和隐私保护、数据质量和准确性、数据集成和互通性、数据分析和洞察、数据可视化和交互性、数据治理和合规性、数据创新和探索性。数据产品是供给「用户」提供数据服务的产品。数据产品的关键特征包括：数据驱动、用户导向、持续迭代、数据安全和隐私保护、数据质量和准确性、数据集成和互通性、数据分析和洞察、数据可视化和交互性、数据治理和合规性、数据创新和探索性。

第 1 章

什么是数据产品

- 1.1 数据产品的关注点
- 1.2 什么是数据应用
- 1.3 什么是效率问题
- 1.4 本章小结

同学习其他细分领域的产品经理的工作一样，要深入地学习与数据产品经理这个职位相关的内容，首先应从了解数据产品开始。

在本章中，笔者通过讲解数据产品的定义，使大家明白数据产品同其他“与数据相关的产品”之间的根本差别。并从数据产品的定义中，剖析数据产品的关注点，以及后续的产品工作的重点和方向，以此来指导数据产品经理的相关工作。

首先，笔者从数据产品的关注点入手。关注点的不同决定了将“数据产品”这样一种产品独立出来，并将其作为一种特殊的产品类型配备专业的产品经理、分析师、研发团队等资源的原因。

之后，笔者将从两个方面帮助大家更深入地理解什么是数据产品。第一个方面是数据应用，这个方面的讨论将帮助大家的数据产品形成直观、具体的理解；第二个方面是数据应用中的效率问题，这个方面相对抽象一些，但因为舍弃了具体的形态，大家可以从更抽象、更宏观的视角来看待数据应用中存在的问题，迎接数据产品自身的挑战并抓住机遇。

1.1 数据产品的关注点

截至本书编写时，提到数据产品，大家最直观的认知是与数据相关的产品就可以叫“数据产品”了。但在互联网的世界中，各种产品都或多或少同数据有关，特别是那些偏向中台和后台的、不直接面向终端用户的产品。比如，运营平台、营销平台，以及更加通用化的配置平台、管理平台等。

其中一部分平台产品，甚至已经涵盖了一些基本的数据分析功能。比如，在运营平台中，可以看到新增用户数和活跃用户数。同时随着业务规模的扩大，这些平台也逐步加入了“大数据平台”的行列——需要应对海量数据，需要考虑数据存储和计算性能之类的问题。

难道这些平台就是“数据产品”了吗？大家心目中典型的“数据产品”应当具备以下功能：支持对各种数据的自由处理，最终做成各种美观的图表或数据报告，还可以在业务出现问题的时候及时提醒相关方采取措施等。与这种概念相比，上文中提到的平台似乎还差了一点儿。

因此，这种理解只能算作对数据产品的大致认知，至少还没有偏离“数据”这个核心。但这个定义并不确切，因为这个定义完全不能告诉我们数据产品要做什么、为什么要做和会有什么困难等关键信息，同时还可能给那些刚刚开始接触数据产品的朋友造成误解。

对于数据产品，笔者给出了一个相对合适的定义：**数据产品**，是为了提高数据应用效率而产生的产品，包括平台型产品、系统功能模块、移动端 App 等多种具体形态。

这个定义中有两个词需要解释：“数据应用”和“效率”。

先说“效率”，因为这可能是“数据产品”和“与数据相关的产品”之间的最大差别。效率等同于数据应用过程中的 ROI (Return on Investment, 投资回报率)。常见的数据应用效率问题，包括如何减少人力投入、如何缩短数据处理的时间周期、如何降低出错概率、如何确保准确性、如何提高及时性、如何提高数据的精细度等。这部分内容很重要，笔者在后续章节中还会对其进行展开和细化。

其次是“数据应用”，这个概念相对容易理解。简单来说，数据应用就是使用数据的过程。从数据处理的角度来讲，其内容包括数据的采集、存储、计算、可视化、消息触达等；从数据治理的角度来讲，其内容包括集成、治理、质量保证等；从应用目的的角度来讲，其内容包括统计、监控、分析、洞察，以及基于数据的决策等。这些都属于数据应用的具体场景。

那么这两个词拼起来，为何就能定义“数据产品”了呢？

举个简单的例子，在没有数据产品之前，做一份数据分析报告，需要一位分析师手工完成从数据提取、计算、可视化，到分析、总结、报告撰写的整个流程，耗时 3 个工作日。而有了数据产品之后，提取、计算和可视化的过程，可以依据事先配置的任务自动完成，甚至还可以标注出需要特别关注的关键信息点，整个流程不超过 10 分钟。因此，在数据产品的帮助下，一位分析师做一份数据分析报告的时间周期，从 3 个工作日缩短为 0.5 个工作日。这个过程就被称为数据应用效率的提升，是数据产品存在的意义和价值。

上文中提到的那些产品，虽然也会考虑到效率问题，但并不会将数据应用的效率作为最核心的考虑因素。它们关注的核心点，可能是执行运营策略的便利性（效率）、改变活动策略的便利性（效率）、改变系统配置的便利性（效率）等。

当然，在数据产品搭建的初期，“有数可用”确实是首要的目标。比如，采用