

《软件和集成电路》杂志社 / 编著



数据新势力

发掘大数据的新生力量

|||||
INNOVATION POWER OF
BIG DATA
|||||



中国工信出版集团



电子工业出版社
<http://www.phei.com.cn>

《软件和集成电路》杂志社 / 编著



数据新势力

发掘大数据的新生力量

||||| INNOVATION POWER OF
BIG DATA |||||

电子工业出版社
Publishing House of Electronics Industry
北京 • BEIJING

未经许可，不得以任何方式复制或抄袭本书的部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

数据新势力：发掘大数据的新生力量/《软件和集成电路》杂志社编著. —北京：电子工业出版社，2016.12

ISBN 978-7-121-30168-1

I. ①数… II. ①软… III. ①数据处理—高技术产业—经济发展—研究—中国
IV. ①F279.244.4

中国版本图书馆CIP数据核字（2016）第252900号

策划编辑：张 冉（zhangran@phei.com.cn）

责任编辑：张 冉

特约编辑：徐学锋

印 刷：三河市华成印务有限公司

装 订：三河市华成印务有限公司

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：12.5 字数：222千字

版 次：2016 年 12 月第 1 版

印 次：2016 年 12 月第 1 次印刷

定 价：49.00 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zlts@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）88254210，influence@phei.com.cn，微信：yingxianglibook

编 委 会

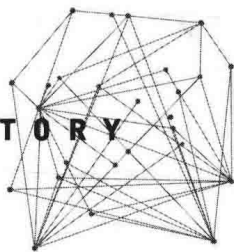
主 编 郭嘉凯

副主编 李 坤

编 委 都莉楠 程梦瑶 张贝贝 张 冉
邱博文 王 盈 郑善双 王 云

INTRODUCTORY

序



自2015年8月，国务院发布《国务院关于印发促进大数据发展行动纲要的通知》以来，大数据产业在国内进入了快速发展期。可以看到，无论是各地政府还是行业用户，对于大数据的重视程度明显增强，同时，大数据与传统行业融合的成功案例也不断涌现。虽然，目前还只是行业内的少数企业在应用，但通过这些领先企业的示范和带头作用，他们应用大数据所取得的成功经验很快就会被行业内的其他企业所效仿，进而带动整个产业的发展 and 壮大。

而作为一个新兴的产业，大数据产业的发展无疑需要更多创新的技术和商业模式作为支撑。同时，国内用户对于大数据的巨大需求空间，也为更多创新型大数据企业的发展提供了坚实的基础和保障。

可以看到，目前国内已经涌现了一批创新型的大数据企业，它们成为推动国内大数据产业发展十分重要的力量。

2016年8月，中国大数据产业生态联盟宣告成立。众多中国优秀的大数据产业链各环节企业、科研院所和行业协会、投资机构，以及地方产业发展机构积极响应，联合发起。

在这些企业中，既有传统IT企业成功转型为大数据企业的代表，也有应产业发展趋势，新近创业的新生代大数据企业。虽然发展经历、规模大小各不相同，但在这些企业身上都能看到一个共同点：对大数据产业的高昂热情，以及对新技术、新商业模式的深入探索。也正因如此，他们才能在众多的企业中脱颖而出，成为中国

大数据产业发展过程中的新势力。

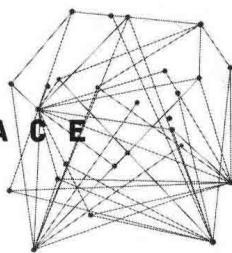
显然，这些企业的转型、创业以及发展经验，对于其他企业都有很好的借鉴意义。因此，在本书中，我们从这些企业中选取了部分典型企业进行了深入的采访。为了更好地展现大数据产业链上各环节企业的全貌，在企业选择方面，我们进行了仔细的甄选：包括行业大数据应用服务商、大数据平台提供商、底层基础架构厂商、咨询服务商等各个细分领域的企业。我们希望通过对这些企业的报道，不仅能够展现中国大数据产业的蓬勃发展现状，同时也能让更多人看到，中国大数据的新兴势力已经崛起！

《软件和集成电路》杂志社总编辑

A stylized handwritten signature in black ink, reading '郭嘉琪' (Gao Jiaqi).

P R E F A C E

前 言



大数据的价值、工具 and 理论

研究和发展大数据，有三个问题特别重要，这三个问题可以总结为六个字——价值、工具、理论。

首先是价值。大数据之所以为各方所重视，是因为它具有价值，是因为它能为我们解决问题提供新的路径。分析大数据的价值，有三类不同的价值链：已有数据集合挖掘新价值、围绕问题求解寻找数据价值、在大数据产业链中提供相关的产品或服务。

一是对已有数据集合挖掘新价值，这是今天大多数称之为“大数据”的企业和机构的主要做法。这条路径的特征是对已有信息集合进行逻辑分析，找到新的结论，从而导向新的价值增长点。

二是针对一些特定问题，特别是对经济社会发展或民生有重大影响的问题，寻找基于大数据的新的解决方案。这是大数据价值三种类型中最重要的部分。大数据的价值，说到底就体现在能否解决我们面临的重大问题上。通常这类价值的实现，起点在对问题的分析，根据问题求解本身的需求，从需要什么信息、如何获取、如何结构化、如何通过特定的数据分析工具与问题相关的应用系统连接。

三是服务于前述两个价值链实现的一组工具。这些工具对大数据实现其价值有帮助，从感知、获取等信息收集到信息的组织管理、结构化、数据特征分析、模型和算法到相关的软件或系统，是一个类型众多，并在不断发展的工具体系。

在互联网上，用大数据这个词搜索，最多的信息是这些工具的介绍或培训相关

内容，其次是基于互联网的信息如何神奇地与一些实际问题相关，但对第二个价值链的关注、分析不够，这个问题值得重视。只要我们认真地分析经济发展、社会管理、公共服务、民生面临的重大问题，都会发现解决这个问题不是已经有足够的数据或信息，缺乏模型、工具、方法，而是没有足以支持问题解决的信息，缺乏高质量的信息是矛盾的主要方面。大数据的价值是围绕着“解决问题”，如交通问题的解决、信用问题的解决、宏观问题态势分析的解决，即数据围绕问题形成价值链。

其次是工具。数据类型的增加、数量的激增、经济社会问题范围的拓展，使得工具日益重要。必须提升工具的能力，使之能与需要解决的数据和相关问题的复杂性匹配。工具可以分成四类：数据的组织管理、与应用系统连接、与数值型数据连接的工具、工具的工具。

第一类是数据组织与管理的工具。30年前，数值型数据库处于发展期，今天数值型数据库管理系统的工具已十分完善，基于这类工具的应用也已十分完美。通过数据字典、数据模型、元数据等方法表述了数据的含义及相互关系，通过规范的数据结构和函数实现数据的有效操纵。那么，今天关于大数据的工具，处理文本、音视频这些类型，应对互联网、互联网条件下的巨大数量，赛博物理空间（CPS）带来的系统复杂性是否能够达到关系数据库管理系统这样的工具水平？回答应该是“相差甚远”。差在什么地方？核心是数据的结构化程度和语义层面的应用。

提升结构化程度，需要对非结构化的信息找到实现结构的方法和规则，要确定颗粒度和应用连接的语义关系。规则、颗粒度、语义关系要有恰当的工具进行处理，保证其有效性，计算的或成本的。需要对数据进行质量和利用的管理，按照确定的质量指标和应用目的，在应用系统中实现其价值。今天，这样的工具不能说没有，但是还没有形成体系，与数据本身的语义逻辑还有很大的差距。

关于信息的结构化，通常有两条路径：一是根据数据本身语义，这是第一类工具的任务；二是根据应用需求，这就是第二类工具。在这两类中，当前最重要的是第二类。一个数据集合，没有应用约束的内在语义关系是极其复杂的，而其价值则是在具有实际意义的应用上。围绕应用需求，把相关数据集合的语义结构揭示出来，形式关系按处理有效性揭示出来，是第一类结构化的前提。围绕应用，要将特定数据集合按形式和语义结构化，首先要确定单元和标识，让数据按照定义的单元和标识进行结构化，其次要根据数据集合和应用需求确定形成单元和标识算法、函数，最后要有相应

的管理系统实现与应用系统一致的管理功能，所以应该是一个工具系列。

第三类是将文本、音视频等非数值型数据管理系统与数值型数据管理系统结合起来的工具。从理论和实践的角度看，到今天为止，技术最成熟、用得最好的还是数值型数据管理系统，非数值型的信息管理系统需要将这两者融合起来。同时，不少应用需要将两类数据同时利用，需要将数值型数据库和非数值型数据库为解决同一个问题服务。更加值得注意的是，当非数值型的数据管理工具成熟时，或者说在形式和语义两个层面充分结构化时，将呈现与数值型数据管理系统雷同的处理需求，这就是第三类工具的必要性和重要性所在。

第四类是工具的工具，即为了开发上述三类工具，甚至包括第四类工具，需要提高开发效率和质量的工具，这是软件工程发展中已经确定的规律。我们要根据各类工具的特征，从中找出需要工具支持的部分。从生态链的角度看，还要为用户提供工具，要为客户系统提供服务，我们需要用工具实现更加商业化的管理和服务。

开发工具的工具，需要回到问题的本源，也需要系统地思考：到底需要什么样的工具？这些工具如何形成？这些工具如何改造和使用？而不是仅把目光盯在已有的工具上。这就隐含着另一个问题：已有的工具所形成的思维方法、演进路径，与解决大数据问题需要的工具的路径不对时，怎么办？所以需要在更高的层面考虑，如何根据事物的本质去利用和创新工具。

第三个问题是理论，也就是说大数据在整个信息技术、信息经济、信息社会中的位置和属性究竟是什么？

首先是定义，一个科学的定义是理论的基础。对大数据有很多定义，也有很多关于大数据特征的讨论，但有一个结论是清晰的，没有一个被称为大数据的客体不是信息，当然，称为信息的客体不一定是大数据。在从IT时代到新的IT时代、从IT时代到DT时代、从工业社会到信息社会、从工业经济到信息经济的发展过程中，不变的是信息。从这样的角度看，大数据是“信息”这个词在当前特定环境下的一个代表性称谓。这几年，互联网和物联网导致非数值型数据爆炸式增长和新价值的发现，大数据成为信息这个一般性名词在当前阶段特殊性的有效指代。我们在研究大数据及其相关的技术系列和社会演进过程中的定位和属性时，应该把它看作信息在当前阶段的一种称呼和标签，不能将它与和信息相关的研究区分开来，尤其不能离

开信息的基本属性。

其次，大数据具有经济属性，实际上这是信息具有经济属性的缩小版。今天经济发展理论面临着一个重要的变量——信息成为整个经济活动过程中的要素，信息自身成为可以独立于传统产品并具有使用价值和交换价值的新产品。这样的要素和产品，与原来的以材料、能源为基础的要素和产品具有不同的属性：它不遵循材料、能源的基本规律，这才有了零边际成本、共享经济等一系列经济模式。所以，大数据（信息）具有特殊的经济属性。

最后是大数据（信息）的社会属性。信息的获取、处理和利用模式的改变，也改变着社会发展中的每个自然成员和机构成员，每个功能系统的发展轨迹。在这一点上，我们虽然已经开始研究，但需要加大研究的力度，提升研究的前瞻性和战略性。

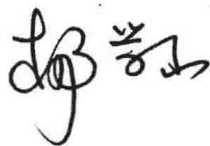
我刚才讲了价值、工具和理论，与其说提出了一些看法，不如说是希望引起大家对这三个方面的重视。其实我们讲价值时，主要是讲商业模式的价值。商业模式是什么东西？你发现价值，并让这个价值实现，让获取价值的人和你本人在其中获利。大数据也是这样，所以价值是分析的基础。

我讲工具其实是在讲技术创新。中国人多，经济发展速度快、规模大。因此，当万物互联或互联网与所有社会构成因素互联时，我们天然拥有数据及其应用的优势，而这种优势为什么不能转变为我们科技创新的优势，成为工具创新的优势呢？

我们再回顾工业革命以来的历史，可以看到领先的国家首先具有高端市场和先进的技术。中国要实现第三个现代化战略目标，必须尽快走向领先的技术和领先的市场。在这一轮与大数据相关的变革中，中国有什么理由不因此造就领先的市场和技术呢？

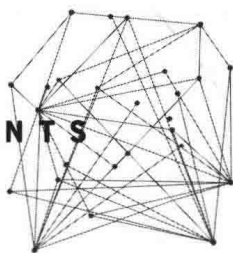
我讲理论，实际上是讲理论创新。当中国从跟随者走向并肩者、领先者时，当中国真正进入发达国家的第一阵营时，如果没有理论创新那是说不过去的。所以我们需要加强理论创新。当我们的理论创新真正取得世界级成果时，我相信中华民族的伟大复兴就为之不远了！

北京大学教授、工业和信息化部原副部长



CONTENTS

目 录



PART ONE | 第一部分

2016中国大数据产业生态地图

第1章 中国大数据产业生态发展概况 / 002

第2章 区域大数据发展与应用 / 016

第3章 中国大数据发展趋势展望 / 021

PART TWO | 第二部分

数据新势力

第4章 久其软件：内生外延 洞悉大数据之生态构建 / 026

第5章 美林数据：大数据推动中国制造业升级 / 032

第6章 企服中国：大数据“云加端”助力产业服务开启新篇章 / 038

第7章 神州信息：大数据全面布局 领跑多行业 / 042

第8章 荣之联：大象无形 数据无极 / 046

第9章 软通动力：数据驱动智慧城市创新发展 / 050

第10章 数据堂：“数据银行”开启大数据破冰之旅 / 055

第11章 百分点：智能重构数据世界 / 060

第12章 普元信息：数据治理开启大数据“淘金”之旅 / 065

- 第13章 神州泰岳：文本技术承载商业模式转型 / 070
- 第14章 星光数据：第一时间挖掘全量数据 / 076
- 第15章 人大金仓：从传统数据库到新兴大数据 / 082
- 第16章 华院分析：企业精准营销创造新未来 / 087
- 第17章 明朝万达：坐在安全之巅方能静看云卷云舒 / 092
- 第18章 昆仑数据：制造即服务 数据即价值 / 098
- 第19章 ZETTAKIT：超融合让云计算回归本质 / 104
- 第20章 中奥科技：聚焦公安行业 深挖数据价值 / 109
- 第21章 R7：数据智能融合开启完美大数据时代 / 114
- 第22章 AMT企源科技：打造中国咨询服务业的“优步” / 119
- 第23章 数联铭品：大数据护航BBD助力金融风险管控 / 124
- 第24章 邦盛金融：破冰金融风险 大数据保驾护航 / 128
- 第25章 科技谷：掘金交通出行大数据 / 132
- 第26章 云房数据：“楼盘字典”让房屋“立”起来 / 137
- 第27章 网思科平：传感器+平台 重构数据应用模式 / 141
- 第28章 三盟科技：“智慧教育”从云端落地 / 148
- 第29章 勤智数码：大数据需要一点工匠精神 / 154
- 第30章 天玑科技：大数据从基础设施起步 / 159
- 第31章 东方金信：迎难而上 聚焦大数据平台 / 163
- 第32章 网智天元：构建影视新生态 需借大数据东风 / 168
- 第33章 朗坤智慧：构建工业企业大数据平台 / 174

跋一 / 180

跋二 / 183

跋三 / 185

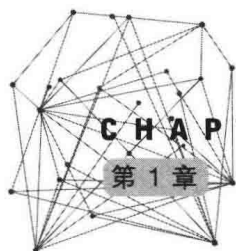
致谢 / 188

PART ONE

第一部分



2016 中国大数据产业生态地图



CHAPTER 1

第 1 章

中国大数据产业生态发展概况

—— 大数据产业分类 ——



融合应用产业
在业务应用中产生大数据，并与行业资源相结合开展商业经营的企业



基础支撑产业
提供直接应用于大数据处理相关的软硬件、解决方案及其他工具的企业



数据服务产业
以大数据为核心资源，以大数据应用服务为主业开展商业经营的企业

融合应用



产业

- 百度
- 阿里
- 腾讯
- 中国电信
- 中国移动
- 中国联通

基础支撑



产业

- 人大金仓
- 浪潮
- 华为
- 南大通用
- TalkingData

数据服务



产业

- 美林数据
- 星环科技
- 数据堂
- 百分点
- TalkingData

大数据产业涉及领域广泛，产业边界尚不清晰。为便于理解、统计和分析，本报告提出了大数据产业的三个口径。

狭义的大数据产业

指数据服务产业，主要由以大数据为核心资源、以大数据应用服务为主业而开展商业经营的企业。

扩展的大数据产业

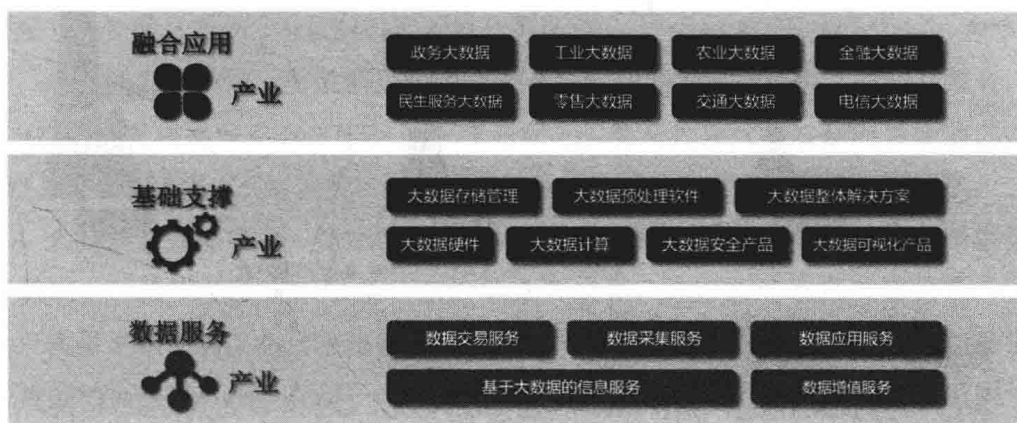
在狭义的大数据产业概念上加上基础支撑产业，即提供直接应用于大数据处理相关软硬件、解决方案及其他工具的企业。

广义的大数据产业

在扩展的大数据产业概念上加上融合应用产业，即各行各业在业务应用中产生大数据，并与其行业资源相结合而开展商业经营的企业。

这三个定义，只是表明其不同的业务特点，也是为了统一口径。毫无疑问，这三类产业都是大数据产业生态中不可缺少的内容。

大数据产业链条



数据服务、基础支撑和融合应用这三类大数据产业，相互依托、有机衔接，共同构成大数据产业链。

数据服务产业

包含数据交易服务、数据采集服务、数据应用服务、数据增值服务以及基于大数据的信息服务等以服务为核心的产业链环节。

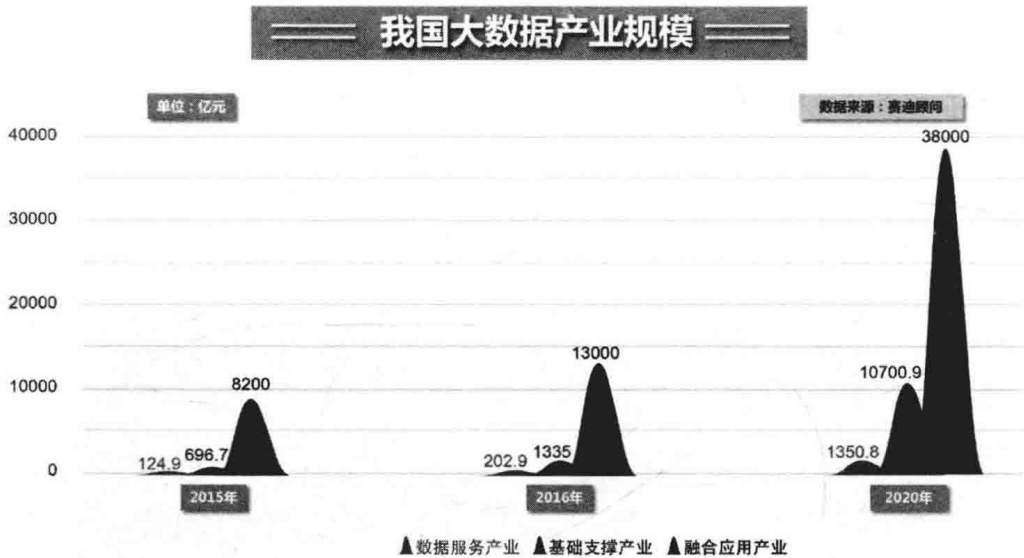
基础支撑产业

包含存储管理、计算处理、可视化、安全等大数据基础软硬件和整体解决方案等产业链环节。

融合应用产业

包含与政务、工业、农业、金融、交通、电信等各行业中与行业资源紧密融合的大数据应用等产业链环节。

中国大数据产业链正在形成过程中，每个关键环节均有一些代表企业。
三个口径的大数据产业，从规模上看差距甚大。



数据服务产业规模

2015年达到124.9亿元，预计2016年达到202.9亿元，2020年将达到1350.8亿元。

基础支撑产业规模

2015年达到696.7亿元，预计2016年达到1335亿元，2020年将超过1万亿元。

融合应用产业规模

2015年达到8200亿元，预计2016年达到13000亿元，2020年将达到38000亿元。
可见，大数据产业的核心是数据服务产业，经济效益体现在融合应用产业。



大数据企业的基本特征是具备大数据技术能力。大数据技术主要体现在大数据采集、预处理、存储、分析挖掘和可视化等环节。

大数据采集技术：有数据采集硬件技术和软件技术两大类，二者交互融合，通过IOT端口对现实世界有关数据进行高速度、大容量长期收集，在网络端通过各类API、网络众包等方式对消费者使用习惯等数据进行高速度、大容量长期收集。

大数据预处理技术：包括数据清理、集成、转换、归约等技术，将海量数据进行标准化、规则化处理，以便于高效率分析挖掘。数据清理就是对数据格式标准化，清除异常数据，纠正错误数据，清除重复数据。数据集成就是将多个数据源中的数据结合起来并统一存储，建立数据仓库。数据转换就是将数据转换成适用于数据挖掘的形式。数据归约就是根据目标适当缩减数据规模，在尽可能保持数据原貌的前提下最大限度地精简数据量，以便提高数据分析挖掘效率。

大数据存储技术：针对大量非结构和半结构化数据以及实时化分析要求而发展起来的高速度、大容量存取数据的技术，包括分布式Hadoop架构、数据库体系和扩展架构。主流为NoSQL和NewSQL的数据库体系、兼容批处理和实时分析的Lambda架构。