

# 大数据行业概况及竞争格局

张君子

## 一、行业概况

### 1、基本情况

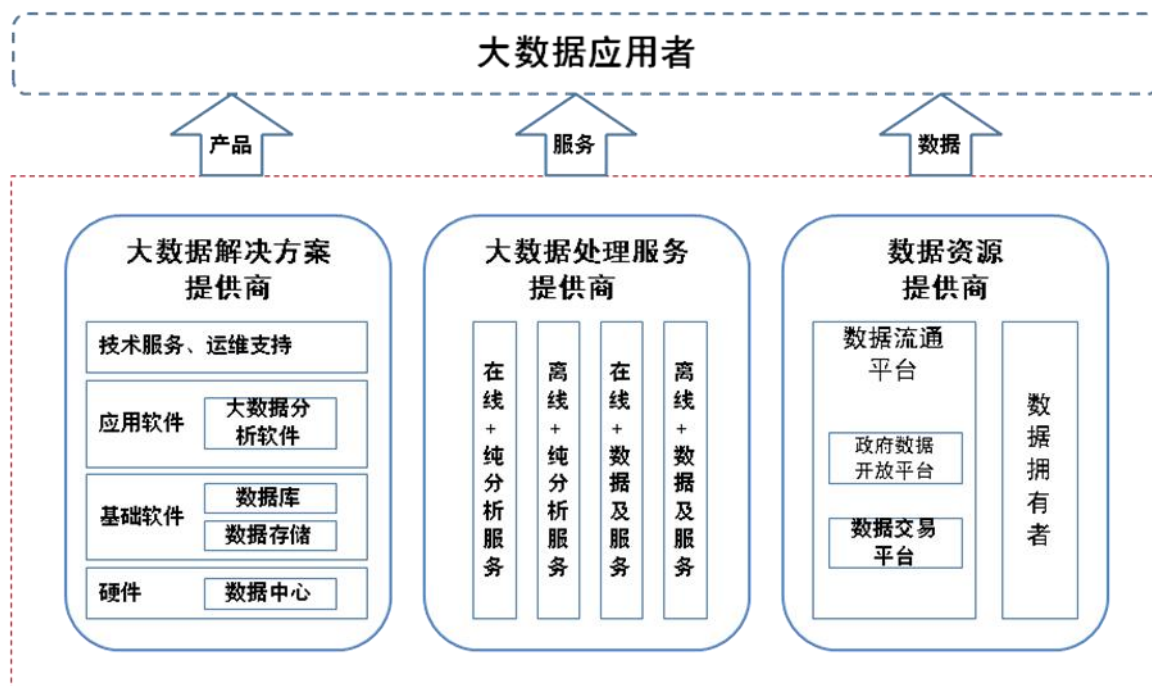
大数据是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。大数据核心价值不在于庞大的数据信息，而在于对海量数据进行筛选、处理，并最终实现数据价值发现和增值的过程。

大数据的特点体现在四个维度：

- ①数据量大：计量单位可达到 PB（1000TB），乃至 ZB(10 亿个 TB)。
- ②类型繁多：包括结构化、半结构化、非机构化数据。
- ③价值密度低：有价值信息被淹没在海量信息中。
- ④速度快、时效高：处理速度要跟上数据增长速度，时效性要求高。

大数据对未来工业互联网、各国经济发展有着至关重要的作用，因此，欧美等发达国家都极为重视大数据的发展。欧美等发达国家早在 2009 年就已经开始从政策层面推动大数据的发展。据麦肯锡预测，预计到 2020 年，大数据可带动美国 GDP 提升 2%至 4%，即创造 3800 亿至 6900 亿美元的价值，并创造 170 万个新的工作岗位。中国大数据发展虽然政策起步较晚，但从 2015 年以来已相继发布了三个和大数据、云计算相关的国家级政策，充分显示了中国对未来大数据、云计算的重视程度。

对比国内国外的大数据生态圈，我们发现国外的大数据企业主要聚集在大数据产业链中的数据源、跨平台、基础设施类、分析类、应用类等方面。其中，相对成熟的公司主要专注于基础设施类，而新创公司主要聚焦于数据源、分析、应用等领域。反观国内生态圈，国内大数据厂商则集中在应用和分析领域，这说明国内公司更愿意直接面对消费者及市场。



大数据产业链可以分为四个部分：数据采集和整合、数据存储和运算、数据分析和挖掘、数据应用和消费。通过技术手段从互联网、移动终端、物联网、应用软件等采集数据，然后把数据按照一定的规则进行存储和运算，再按照需求调用数据并进行智能分析和挖掘，将数据转化成价值信息或者产品，为决策支持、提升效率、创新产品提供依据。

大数据行业的上游行业主要由软硬件供应商构成，国内大数据行业的主要硬件供应商为国外厂商。大数据软件平台主要由基础硬件和工具平台类软件组成。上游行业市场内部竞争较为充分，行业格局比较稳定，呈现价格稳中趋降、性价比逐年提高的趋势。

大数据行业的下游行业主要为终端使用客户，主要包括电信、金融、能源、交通和政府等信息化程度依赖相对较高的社会公共服务事业领域。随着企业对大数据接受程度的提高，大数据行业的下游覆盖面正在日益拓宽。

## 2、行业发展趋势

### ①行业继续保持快速增长

从全球市场看，大数据产业近年来一直保持较快的增长势头，根据 Wikibon 数据显示，2014 年全球大数据市场规模达到 285 亿美元，同比增长 53.2%，预计 2015-2017 年全球大数据的市场规模分别可达到 383 亿美元、452

亿美元和 500 亿美元。而据易观国际数据显示，2014 年我国大数据市场规模达到 75.7 亿元，同比增长 28.4%，预计 2015 和 2016 年将分别达到 98.9 亿元和 129.3 亿元，两年的同比增速均为 30.7%。

## ②政府大数据迎来开放

目前我国政府各部门的数据相互封闭，不仅制约了行政服务效率，而且制约了大数据产业的发展。2015 年 8 月 31 日，国务院印发《促进大数据发展行动纲要》，国家层面的大数据战略思路逐渐成型并发布。《促进大数据发展行动纲要》将大数据战略定位为推动经济转型发展的新动力、重塑国家竞争优势的新机遇、提升政府治理能力的新途径；提出 2017 年底前形成跨部门数据资源共享共用格局、2018 年底前建成国家政府数据统一开放平台、2020 年底前逐步实现信用、交通、医疗、卫生、就业、社保、地理、文化、教育、科技、资源、农业、环境、安监、金融、质量、统计、气象、海洋、企业登记监管等民生保障服务相关领域的政府数据集向社会开放。

政府数据的对外开放将为大数据行业带来宝贵的数据资产“活水”，将带动整个行业快速发展，同时使沉睡在政府手中的数据实现价值变现；而政府数据的内部共享将打通各部门的数据间隔，在外部技术合作伙伴的协助下实现数据融合，并在此基础上进行数据分析，提升政府治理能力，同时令大数据技术合作方充分受益。

## ③大数据与产业结合更加紧密

中国近年来大数据发展迅速，得到广泛关注，在走过探索期、市场启动期，当前大数据产业已经在接受度、技术、应用等各方面趋于成熟，大数据在企业中应用的巨大价值逐渐显现，也开始被各行各业所认识和接受。随着大数据的高速发展，其渗入线上线下行业的应用也逐渐走向成熟，行业应用的增加也正有力推动整个行业的市场规模的快速增长。下图表示大数据在不同行业中的应用场景：

行业	应用示例	行业	应用示例
电子商务和 在线零售	提高客户终生价值	媒体	有线行业，大型有线运营商每天都可以使用大数据来分析机顶盒数据，可以利用此数据来调整广告或促销活动
	提供一致的跨渠道客户体验		
	从销售、营销和其他来源收获客户线索		
	持续优化后端流程		
金融服务	合规性和监管报告	欺诈检测	信用卡和借记卡欺诈
	风险分析和管理		存款帐户欺诈
	欺诈检测和安全分析		技术欺诈和坏账
	CRM 和客户忠诚度计划		医疗欺诈
	信用风险、评分和分析		医疗补助计划和医疗保
	高速套利交易		财产和灾害保险欺诈
	交易监管		工伤赔偿欺诈
异常交易模式分析	保险欺诈		
健康和生命 科学	健康保险欺诈检测	电信	收入保障和价格优化
	营销活动和销售计划优化		客户流失预防
	品牌管理		营销活动管理和客户忠
	患者护理质量和程序分析		呼叫详细记录(CDR)分析
	医疗设备和药物供应链管理		网络性能和优化
	药品发现和开发分析		移动用户位路分析

#### ④大数据与云计算进一步融合

由于企业在应用大数据分析时会面对诸多新的复杂的困难与挑战，云计算服务成为大数据分析的一个理想的载体，而大数据分析服务也能提升云计算平台的市场竞争力，提升云计算的附加价值。从技术上看，大数据和云计算的关系是相辅相成的，大数据必须采用分布式架构，在对海量数据进行分布式数据挖掘的过程中，大数据必须依托云计算的分布式处理、分布式数据库和云存储、虚拟化技术。

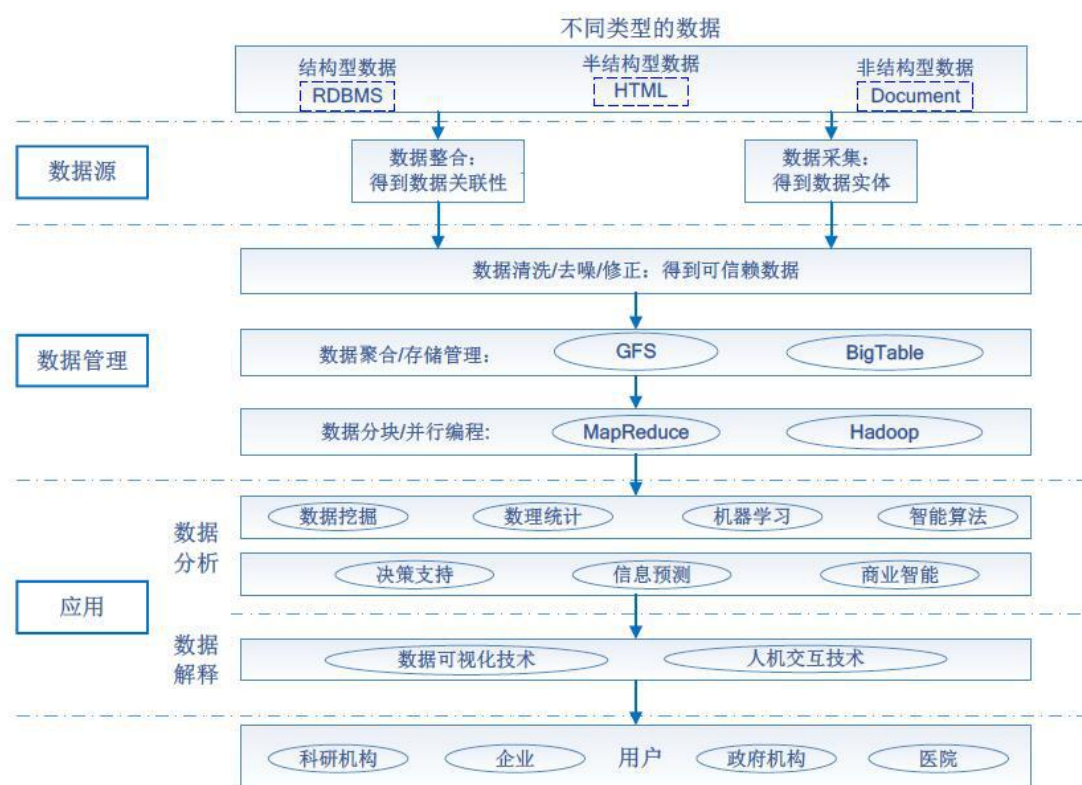
大数据分析对计算和存储硬件需求高，需求随时多变，既可能有大段低需求时间，也可能有突发巨量计算资源需求；而云计算服务可按需求计费，不需要一次投入大量资本建设一个可能有闲置的计算中心，当需求增长时可快速部署新的计算资源服务。

大数据分析需要特殊的数据管理软件应对复杂的数据源、快速变化的数据结构、不断伸缩的数据总量，云计算存储结构天生适合处理复杂多变的数据源，云计算服务商通常提供现成的复杂数据库存储管理软件。

### 3、大数据技术

大量的新技术被开发出来并应用在大数据处理生命周期的三个环节：数据

源，数据管理，以及应用。



数据来源：浙江大学学报《大数据技术研究综述》

### ①大数据源：数据库技术

大数据与传统数据最重要的区别即是数据类别由原来的结构化数据，加入了半结构/非结构化的数据。为解决半结构化/非结构化数据的爆发，NoSQL 数据库应运而生。

### ②大数据管理技术：分布式架构

大数据计算是数据密集型计算，对计算单元和存储段元间的数据吞吐率要求极高，对性价比和扩展性要求也非常高，因而分布式计算被大规模应用到了大数据领域。2004年，Google 公布了 MapReduce 分布式并行编程架构；而后，Yahoo 提出 S4 系统、Twitter 提出 Storm 系统，以实施“边到达边计算”的实时流计算框架；Google 随后提出了将 MapReduce 内存化以提高实时性的 Spark。Apache 基金会下开源分布式计算平台 Hadoop 以分布式文件系统 HDFS 和 MapReduce 分布式计算框架为核心，为用户提供底层细节透明的分布式基础设施。

### ③大数据应用：分析与解释

数据分析与解释横跨了计算机科学、统计学和数学、专业领域知识等多个领域。近年来，传统的数学和统计学数据分析越来越难以满足大数据时代的需求，机器学习和人工智能渐渐流行。

## 4、影响行业发展的有利和不利因素

### (1) 有利因素

#### ①国家政策的大力支持

政府拥有最多且最具应用价值的核心数据，一方面大数据已经上升至各国国家战略，政府推进大数据开放已经是大势所趋：美国等发达国家纷纷布局大数据产业，推出大数据相关政策。大数据对整个世界的的影响力呈现快速增长趋势，将引发新一轮大国竞争。另一方面我国政府的大数据政策方向已经非常明确：政府公共部门数据整合与开放，以及大数据技术研发与产品化。近年来，大数据多次出现在政府部门工作会议、政策支持文件及十三五规划文件中，可见，在未来的一段时间内，国内大数据领域将迎来投资、建设高潮。

#### ②云计算助推大数据产业发展

云计算已经从概念走向实际应用，已经进入高速发展期。根据 Gartner 数据显示，2012、2013、2014 年全球云服务市场规模分别达到 1110、1310、1528 亿美元，分别同比增长 23.33%、18.02%和 16.64%。全球云服务市场受制于全球整体信息化水平的影响，增速略小于大数据行业。

云计算降低了使用 IT 资源的门槛，为数据集中化创造了基础，极大的促进了大数据产业的发展。未来企业将不用再购买服务器，直接购买终端，输送至数据中心，从而形成服务器集群产业链，满足了大数据存储和挖掘的需求。2015 年 2 月，国务院下发《关于促进云计算创新发展培育信息产业新业态的意见》提到：开展基于云计算的大数据应用示范，支持政府机构和企业创新大数据服务模式，政府部门要加大采购云计算服务的力度等一系列措施。

#### ③相对封闭的蓝海市场

由于大数据领域涉及国家数据/信息安全问题，未来中国这个市场会相对封

闭，外资企业开展业务会受中国政府监管的限制，中国大数据产业，特别数据工具产业是一个对外封闭的蓝海市场。中国特殊的国情甚至能让中国企业拥有全球领先的数据资源并开发相应应用。

## （2）不利因素

### ①缺乏创新意识，研发投入不足

缺乏具有创新意识的企业是目前大数据行业面临的问题，也是国内科技行业的通病。企业研发投入不足导致了国内大数据行业较国外发展缓慢。IBM、EMC等在大数据领域的并购、研发投入都高达百亿美元，才造就了大数据技术的不断发展进步。例如 IBM 在 2014 年投资 240 亿美元构建其大数据与分析能力。当年公司商务分析业务（Business analytics，该部分业务包括大数据）收入增长 7%，至 170 亿美元。

### ②数据所有权不清晰

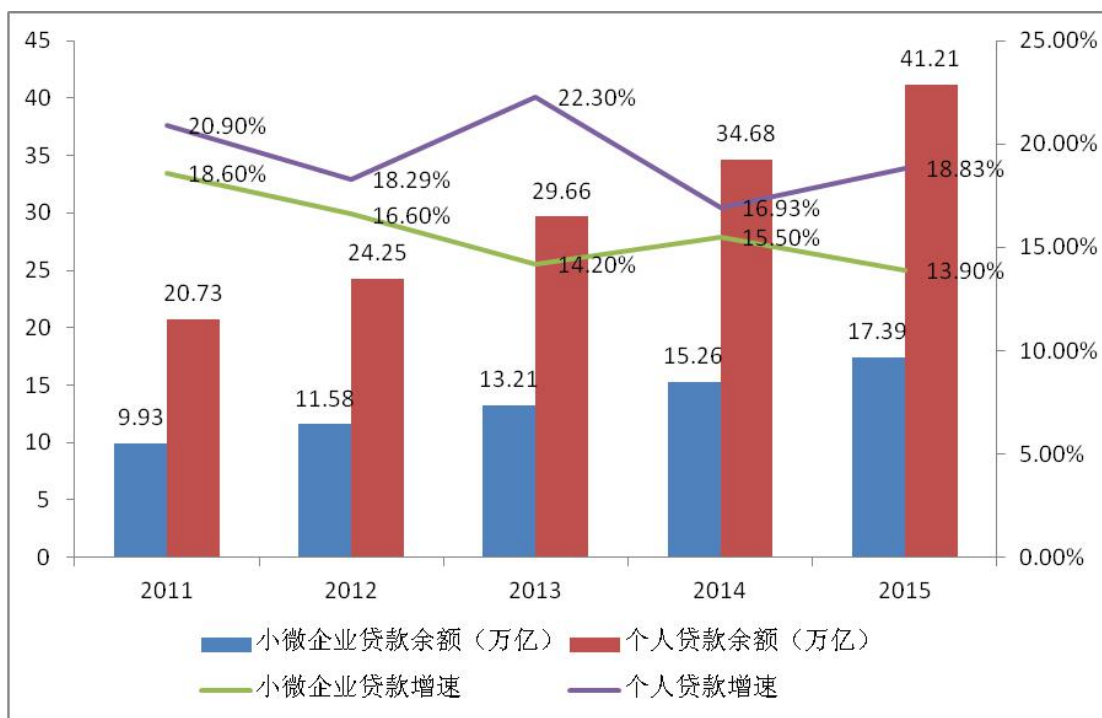
当前数据源公司在开发、使用或者变现数据中，仍然保有对数据开放的谨慎态度、对所有权界定、隐私保护等问题的担忧。而产品类公司，是不可能也不可以独占数据增值的价值，他们属于从属地位。在数据源开放有限的现实背景下，整个大数据产业链下游也将受限于此。

## 5、大数据征信市场

我国的征信行业起步较晚，经过 20 多年的发展，形成了以人民银行征信中心为主导，民营征信机构为补充的混合经营模式。目前人民银行已经建立起覆盖全国的公共征信网络，民营征信机构业务逐步向市场化迈进，整个行业进入快速发展期。

### ①征信市场边界不断扩张

近年来我国小微企业贷款、个人购房贷款、消费性贷款和经营性贷款稳步增加，在需求端为未来个人征信业务发展打下良好基础。2015 年小微企业贷款余额为 17.39 万亿，同比增长 13.90%；个人贷款（含个人购房贷款、消费性贷款、经营性贷款）余额为 41.21 万亿，同比增长 18.83%。



数据来源：中国人民银行

自 2007 年以来，P2P 网贷、第三方支付、互联网保险、虚拟信用卡、网络银行、股权众筹等互联网金融业态相继出现并成熟，尽管许多互联网金融公司和民间机构有强烈的征信数据需求，但并不在央行征信系统服务对象之列。征信作为互联网金融体系中的基础设施，是互联网金融业务风险控制、产品定价的基础，互联网金融健康发展的需求必将，倒逼“基础设施”跟上发展步伐。



②大数据为征信业发展提供技术“新利器”



以大数据为代表的新兴技术为征信产业链注入新活力，大数据可解决海量征信数据的采集和存储问题、通过机器学习和人工智能方法可深入进行征信数据挖掘和风险分析、借助云计算和移动互联网等手段可提高征信服务的便捷性等。与传统征信企业不同，互联网企业的信用数据时效性更强、来源也更为丰富。电商网购、在线支付、信用卡还款、水电煤缴费、社交信息等都可以成为互联网征信的原料。依托于大数据和云计算技术优势，可以挖掘大量数据碎片中的关联性，提炼出有价值的信用信息，为征信的有效进行提供更为丰富的数据原料，征信数据规模越来越大、数据维度越来越广，模型不断迭代优化。与传统征信模式相比，大数据征信拥有全新特征：

	传统征信	大数据征信
数据来源	以银行信用数据为主，来源单一，采集频率低	第三方数据：信贷、司法、水电煤缴费等；互联网数据：电商、理财、社交平台等；用户提交的数据：教育、职业等。可以做到实时采集
数据格式	结构化数据	结构化数据和大量非结构化数据
评价思路	用历史信用记录反映未来信用水平	从海量数据中推断身份特质、性格偏好、经济水平等相对稳定的指标，进而判断信用水平
分析方法	采用线性回归、聚类分析、隐私分局和分类树等方法	机器学习、神经网络、Page Rank算法、RF等大数据处理方法
服务人群	有信贷记录人群	有信贷记录人群和在生活中留下足够痕迹的人
应用场景	借贷等金融领域为主	金融领域，多种生活领域

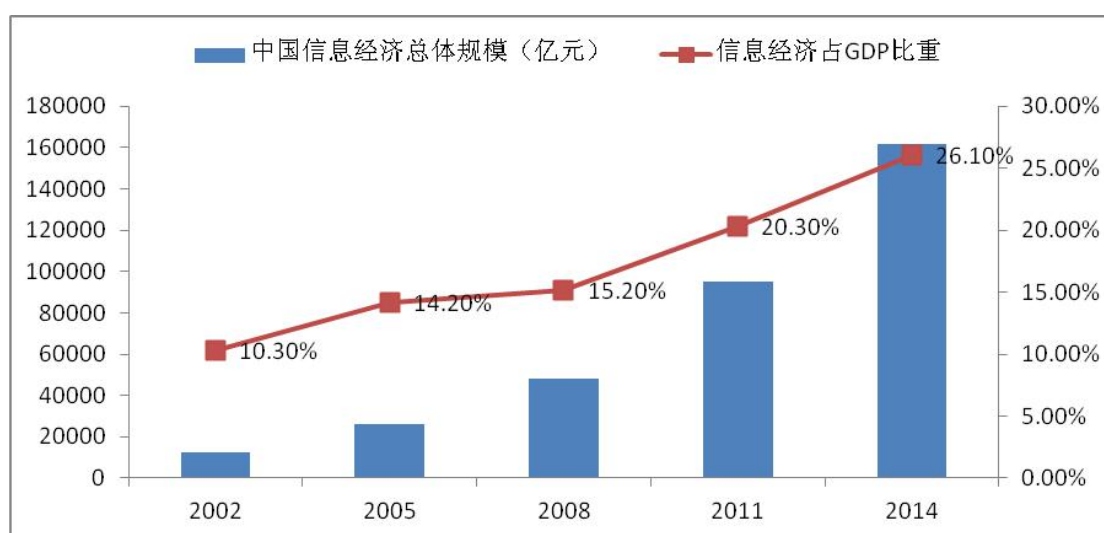
### ③ 征信行业空间预测

2014年美国最大的企业征信机构Dun&Bradstreet营业收入为16.82万亿美元，其中北美地区贡献74%，按市场占有率90%估算，美国企业征信总规模约为13.83亿美元。美国2014年GDP为17.35万亿美元，企业征信收入占GDP比重接近万分之一（约为0.08%），日本的情况近似。2014年中国GDP为63.61万亿元，参考该比例可估算出2014年中国企业征信市场潜在规模约为60亿元。未来中国的GDP增速将保持在7%左右，随着经济总量提升，企业征信规模还将进一步扩大。需求侧，国内企业信贷需求十分旺盛，近年来企业及其他部门贷款余额增速一直保持较高水平，预计随着征信市场化改革的推进，未来国内企业征信年营收规模可达100亿元。我国目前个人征信规模较小，收益分配严重不均，由于主导个人征信业务的央行征信中心并不以盈利为目的，征信报告收费仅为成本价格。随着国内个人征信行业市场化的推进，个人征信机

构将逐渐由非营利性机构转为盈利性企业，信用报告的价格也将大幅提高，届时中国未来个人征信市场长期收入规模将快速增长，业内预计个人征信市场规模在 600 亿元至 1000 亿元之间。

## 二、行业市场规模

据中国信息通信研究院相关数据，2014 年我国信息经济总量达到 16.20 万亿元，同比名义增长超过 21.10%，显著高于当年 GDP 增速，占 GDP 的比重超过 26.00%，同比提升 2.40 个百分点。信息经济已成为近年来带动经济增长的重要动力，目前互联网直接关联产业对我国 GDP 贡献率达 7% 以上，未来仍有较大的增长潜力。

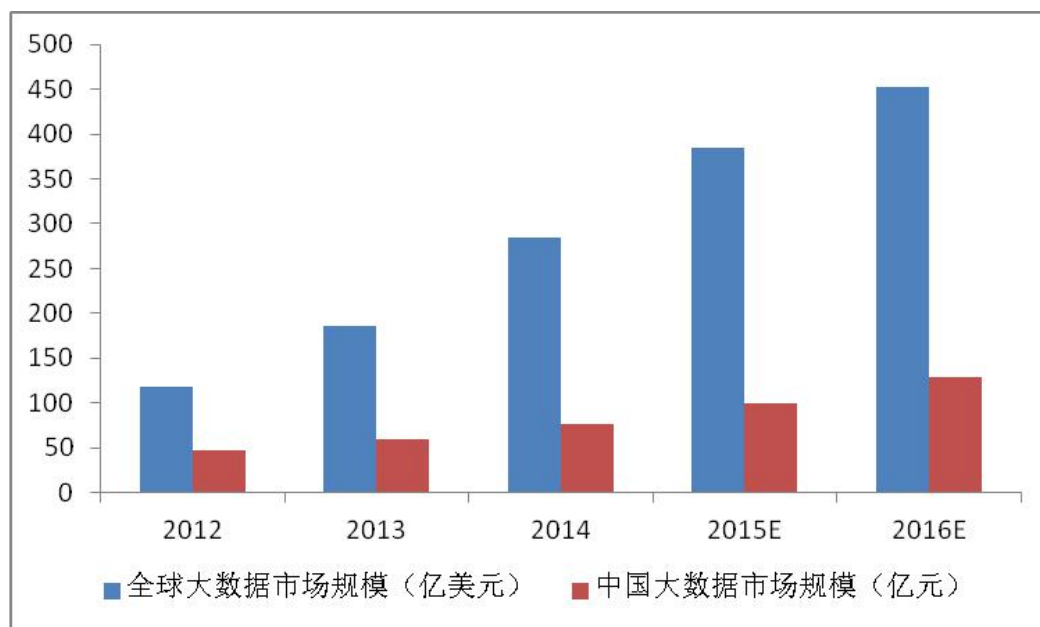


数据来源：中国信息通信研究院

近年来移动互联网和物联网的出现，数据产生的速度和规模开始急剧提升，过去几年产生的数据比以往 4 万年的数据总量还要多。预计随着互联网与产业的加速融合，未来数据体量的增速将呈现指数上升态势。IDC(国际数据公司)预计 2020 年全球数据使用量将达到 40ZB (ZettaBytes)，届时中国产生的数据量将占到全球总量的 21%。

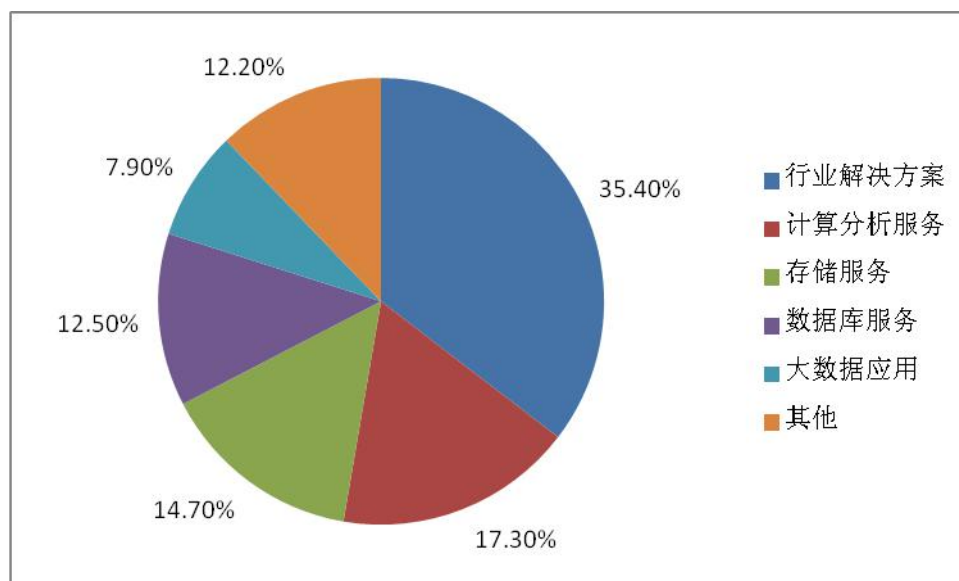
全球市场看，大数据产业近年来一直保持较快的增长势头，根据 Wikibon 数据显示，2014 年全球大数据市场规模达到 285 亿美元，同比增长 53.2%，预计 2015-2017 年全球大数据的市场规模分别可达到 383 亿美元、452 亿美元和 500 亿美元。IDC 在最新的研究报告中指出，预计从 2014 年到 2019 年，全球大

数据市场将以每年 23.1%的复合增速发展，至 2019 年，全球大数据市场规模将达到 486 亿美元。而据易观国际数据显示，2014 年我国大数据市场规模达到 75.7 亿元，同比增长 28.4%，预计 2015 和 2016 年将分别达到 98.9 亿元和 129.3 亿元。



数据来源：Wikibon，易观国际

全球大数据市场中，行业解决方案、计算分析服务、存储服务、数据库服务和大数据应用为市场份额排名最靠前的细分市场，分别占据 35.4%、17.3%、14.7%、12.5%和 7.9%的市场份额。



数据来源：Wikibon

### 三、行业竞争格局

由于大数据在国外起步较早，技术积累较为完备，目前大数据厂商竞争力排名靠前的均为国外知名厂商，其中包括 IBM、Oracle、SAP 等 IT 巨头，依靠其雄厚的资金实力和研发能力，在大数据市场遥遥领先。在国家政策支持、企业内生需求逐步成长的过程中，国内开始涌现不少大数据公司，可以将他们划分为三类：数据源公司、数据管理工具/服务公司、数据分析工具/服务公司。

数据源领域的参与者，典型代表是互联网公司 BAT（百度、阿里巴巴和腾讯）、电信运营商、政府部门和传统行业信息化厂商。数据处理/分析领域参与者，国内已有部分较为专业的独立工具提供商，部分通过转型大数据处理与分析的行业信息化建设厂商。BAT 为代表的互联网企业通过数据资产与数据技术融合发展的方式，推动自身的业务领域从搜索、电子商务、即时通讯、社交网络、广告精准投放等领域逐步向位置服务、智能交通、互联网金融等领域渗透；浪潮、联想、用友为代表的传统 IT 企业正在加速向大数据解决方案提供商转型，并纷纷推出各自的大数据解决方案，其产品和服务多是基于原有业务开展，未能撼动互联网公司的领先地位。新兴的独立大数据工具厂商在大数据处理/分析技术方面，相较其他参与主体具有较强的优势，若能具有较好的数据资源获取能力，可以进一步提升该类公司的竞争力，但是产品化和变现能力仍然对其发展构成挑战。