

大数据行业报告

作者：张千帆

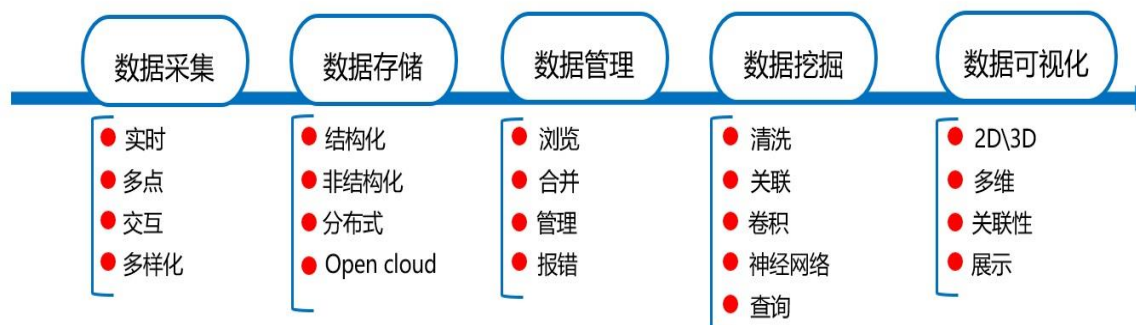
大数据产业是基于当前社会互联网发展中暴涨的数据所衍生出的新兴行业。数据是比语言文字更精确的描述事物本质的工具。数据时代人类的所有记录，无论是机构化的文字、文档还是非结构化的图片、音频、视频，都是以数据的形式存在。

而大数据是计算机和网络科学行业内的术语，从原有的技术语义上理解指的是巨量数据集合。而根据 Wikipedia 的定义“大数据是值在无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”。在维克托（2012）撰写的《大数据时代》中指明大数据是一种数据处理的方式，它与其他传统方式的不同之处是采用所有数据来做分析而不是采用随机分析法（抽样调查）。而研究机构 Gartner 定义大数据是：“需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产”。从以上定义来看大数据的定义包涵两个方面：第一是基于数据的本身和特点，在互联网社会中数据作为一种新的信息资产它自身拥有 4V 特性--：Volume（大量）、Velocity（高速）、Variety（多样）、Value（价值）；第二是对海量数据的处理和挖掘产生的价值，对海量数据本体来说它若未被挖掘处理形成价值，则不具有影响力；而通过数据的采集、存储，传输，处理、挖掘，则能提取出数据中潜在的和有价值的信息。而这个处理的过程和能力才是人们谈论大数据的真正意义。而相比西方研究者强调的预测是大数据的核心，当前的大数据则成为一种思考和行为模式，国内客户观念上的转变远比技术手段的影响要更深远。

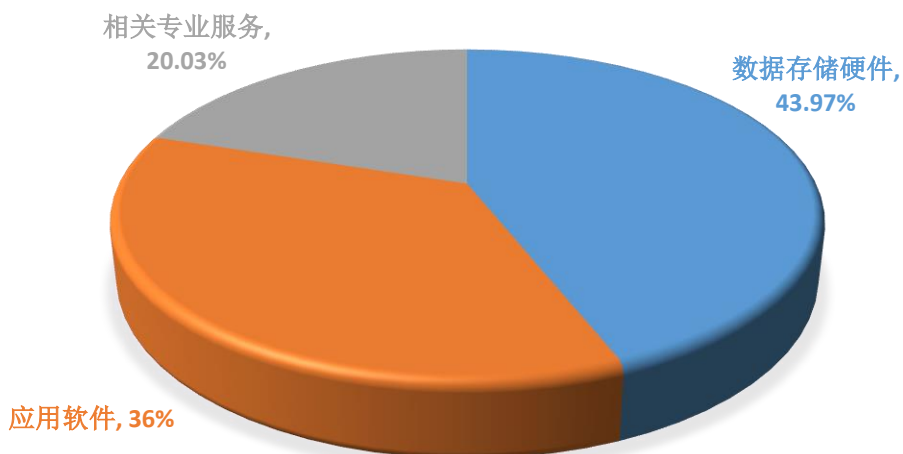
一、 大数据行业发展和国内现状

信息技术的不断演进和移动互联网的跨越式发展导致人类社会产生的海量数据正在以每两年翻番的高速不断增长，而且这个速度将在一段时间内保持不会降低。根据 IDC 的监测统计，2018 年全球范围内数据会达到 20ZB，对于中国来说 2012 年国内数据存储总量就已经达到 364EB，近年来数据更是以指数级爆发增长。如何有效的采集、存储、管理、分析、处理这些数据已经成为传统软件行业所必须面对的挑战。

大数据行业的产业链就如下图所示，存在非常明显的上下游关联关系；也非常明显的对应了行业发展的路径。



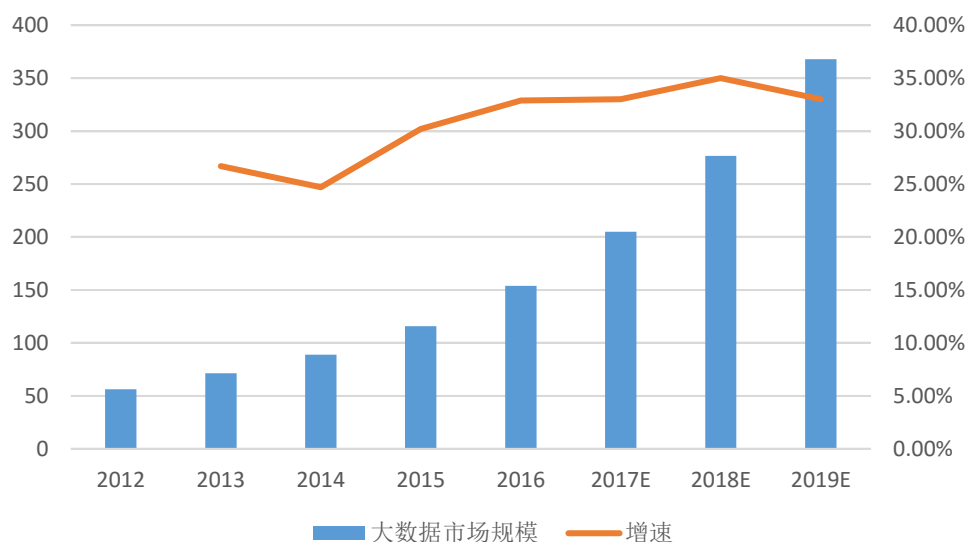
(1) 2009-2013 年我国数据行业处于起步阶段。数据的采集方法和路径已经形成并较为完善，数据的存储能力也稳步提升。这一点可以从大数据行业的市场规模看出来。中国信息通信研究院统计 2015 年国内大数据市场整体规模约为 115.9 亿元，数据存储硬件产值、hadoop 类软件产值以及相关专业服务的产值占比分别为 43.97%、36% 和 20.03%。相比之前 50%-80% 的硬件投入占比，明显可见硬件占比的下降以及应用软件占比的上升。相比发达国家的发展轨迹，后期数据专业服务类占比的将持续提升，并达到三分之一的比例。



2015 年大数据细分市场结构

(2) 2014-2016 年大数据脱离炒作，逐步走向技术落地、向其他行业渗透及推动发展的通道。2014 年由 Gartner 统计的新兴技术发展周期预测的炒作周期曲线上，大数据已经由期望顶点下滑并加速向最低点滑落，2015 年甚至都不再出

现了。这一现象与早期云计算的发展轨迹类似，说明对概念的炒作已经结束，后续进入更细分领域及应用技术的发展周期。我国企业对大数据的潜在价值已经认同，不仅仅认同概念更倾向于落地技术及向生产力平台推进。2015 年的行业市场规模达到 115.9 亿元，增速达 30.19%。而随着各个行业政策的有力推动和示范项目的成功运营，预计 2016 至 2018 年中国大数据市场规模还将维持 30%-35% 左右的高增长。



大数据行业市场规模（单位：亿元）

数据来源：Choice 金融终端

(3) 2016 年是国内大数据发展真正的元年。国内各级政府、各部委及行业都发布了大数据相关的发展规划，仅 2016 年发布的大数据行业重点相关政策就有 10 个。大数据行业不但自身将成长为规模庞大的新兴产业，并有望在 2016-2020 年间，带动市场规模万亿之巨的 IT 服务业转型，促进国民经济其他领域的飞速发展。

序号	发布时间	政策规划名称	发布单位	文件内容
1	2016 年 1 月 7 日	《关于组织实施促进大数据发展重大工程的通知》	国家发改委	将重点支持大数据示范应用、共享开放、基础设施统筹发展，以及数据要素流通。
2	2016 年 3 月 7 日	《生态环境大数据建设总体方案》	环保部	运用大数据、云计算等现代信息技术手段，全面提高生态环境保护综合决策、监管治理和公共服务水平，转变环境管理方式和工作方式。
3	2016 年 6 月	《关于促进和规范健康医	国务院	大力推动政府健康医疗信息系

	24 日	疗大数据应用发展的指导意见》		统和公众健康医疗数据互相融合、开放共享，积极营造促进健康医疗大数据安全规范、创新应用的发展环境。
4	2016 年 7 月 4 日	《关于印发促进国土资源大数据应用发展实施意见》	国土资源部	丰富与完善统一的国土资源数据资源体系。初步建成国土资源数据共享平台和开放平台，实现一定范围的数据共享与开放。
5	2016 年 7 月 8 日	《“大数据的统计学基础与分析方法”重大项目指南》	国家自然科学基金委员会	包括了两项大数据相关重大项目指南：《“大数据的统计学基础与分析方法”重大项目指南》与《“互联网与大数据环境下高端装备制造工程管理理论与方法研究”重大项目指南》
6	2016 年 7 月 13 日	《关于加快中国林业大数据发展的指导意见》	国家林业局	充分发挥林业大数据在生态建设中的重要功能和巨大潜力，推进数据资源开放共享，积极培育林业发展新业态。
7	2016 年 7 月 19 日	《推进煤炭大数据发展指导意见》	中国煤炭工业协会、中国煤炭运销协会	力争在 2020 年建成全国煤炭数据平台，实现煤炭数据资源适度向社会开放，为煤炭行业探索新业态、新模式和行业转型升级提供支撑。
8	2016 年 8 月 26 日	《关于请组织申报大数据领域创新能力建设专项的通知》	国家发改委	重点提出两个方面的专项建设内容，提升大数据基础技术支撑能力为其中一方面。
9	2016 年 8 月 25 日	《关于推进交通运输行业数据资源开放共享的实施意见》	交通运输部	明确了推进交通运输行业数据资源开放共享的实施目标。
10	2016 年 10 月 14 日	《农业农村大数据试点方案》	农业部	贯彻落实有关部署要求，通过试点示范扎实推进农业农村大数据发展和应用。

2016 年 1 月 15 日，贵州省通过了《贵州省大数据发展应用促进条例》，这是中国首部大数据地方法规，以立法推动大数据产业蓬勃发展。条例的出台不仅是贵州作为大数据综合试验区迈出的坚实一步，对大数据产业的健康发展具有很大的促进作用，更为重要的是，条例填补了中国大数据立法的空白。

2016 年 10 月 8 日，国家发展改革委、工业和信息化部、中央网信办发函批复，在京津冀、珠江三角洲、上海市、河南省、重庆市、沈阳市、内蒙古七个区域推进国家大数据综合试验区建设，这是继贵州之后第二批获批建设的国家级大

数据综合试验区。此次批复是贯彻落实国务院《促进大数据发展行动纲要》的重要举措，将在大数据制度创新、公共数据开放共享、大数据创新应用、大数据产业聚集、大数据要素流通、数据中心整合利用、大数据国际交流合作等方面进行试验探索，推动我国大数据创新发展。

2017年1月17日工信部编制并正式印发了《大数据产业发展规划(2016-2020年)》。大数据已经成为国家塑造竞争力的战略制高点之一。《规划》提出了7项重点任务、8大重点工程、5个方面保障措施。目标到2020年，大数据相关产品和服务业务收入突破一万亿元，年均复合增长率保持30%左右；10家核心龙头、500家应用企业、建设10-15个大数据综合实验区等具体指标的制定将使得整个产业从资源开放、基础设施建设以及应用领域等几个层面得到支持，提振产业发展的信心。《规划》全面部署“十三五”(2016-2020年)时期大数据产业发展工作，为实现从数据大国向数据强国转变指明了方向。围绕数据科学理论体系、大数据计算系统与分析、大数据应用模型等领域进行前瞻布局，加强大数据基础研究。以应用为牵引，自主研发和引进吸收并重，加快形成安全可控的大数据产品体系。未来支持电信、互联网、工业、金融、健康、交通等信息化基础好的领域率先开展跨领域、跨行业的大数据应用，支持服务业利用大数据建立品牌、精准营销和定制服务等。

二、 大数据行业发展趋势

在国家宏观经济保持低速增长、信息化进程快速推进、区域经济急需深度提升质量等因素的带动下，对于第一及第二产业的改革已经达到关键点；对于生产力平台提升的需求将继续处于较高水平，基于此我国大数据行业在未来将有望保持良好的发展趋势。

(1) 底层基础硬件及软件市场仍然广阔

随着市场需求的不断提升，越来越多的传统企业意识到数据推动决策和战略的必然性。但是大数据的系统应用必须配合数据云平台的硬件支持，这无疑将加速企业云化的过程。无论是共有云、私有云或是混合云的搭建都直接助长了传统ICT厂商的巨大市场空间。而大数据的底层工具软件包括Hadoop、Spark都是从零开始在十年的时间里完全改变了企业数据平台和解决方案，并逐渐解决了企

业数据存储、处理、分析的各种关键问题。目前，还出现了其他的一些新的框架，比如 Flink、Ignite、Samza、Kudu 等，这些框架的发展也非常迅猛。但从整体上来说，更高速，更易理解、更容易编程、更少载荷的硬件及软件是行业发展的一贯追求，不但现在如此，未来也将持续如此。

(2) 通用性和专业性的双向发展趋势

大数据行业中最被关注的核心是分析及挖掘；与互联网的发展路径相类似，数据将向通用性和专用性两个不同的方向发展。以 Tableau 例的通用性软件充分满足了普通用户的大部分需求，控制台可以直接监测数据，也提供完整的数据分析能力；而且还能够自主配置软件台功能。对于大部分行业客户来说，该类型数据软件的使用门槛低，而且自制模板的网络共享平台也极大程度上满足了普适化的需求。老牌软件公司 IBM、Oracle 等也已经加入到大数据分析的阵营中，不断升级其产品并推动其云化进程。

而另一个发展方向则是针对垂直细分行业提供专业化的分析。英国零售巨头 Tesco 旗下的数据公司 Dunnhumby 就曾在 2005 年基于 Tesco 的客户数据运行过一个“忠诚计划”。通过分析大数据，来跟踪顾客的购物习惯，并建立消费者档案。采集和分析消费数据能够透过消费行为识别顾客需求，划分类别能够摸清这一群组消费者的个性与共性，从而在日常消费中进行精准引导、营销为消费者提供更好的商品服务。该计划使 Tesco 取代 Sainsbury 成为英国最大的连锁超市。Dunnhumby 的分析能力明显是针对零售流通行业最好的“金手指”，但对于 Google 这样的公司来说却并非有决定性作用。该类型的数据挖掘能力需要大量行业数据的积淀，同时也需要不断的优化和调整模型，从而达到最优化的决策产出。针对每个细分行业的数据分析都存在巨大的市场，虽然其发展的速度较通用性软件较慢，但无可否认将持续出现的“独角兽”公司将更具优势。

(3) 行业龙头带动整体发展

互联网厂商作为大数据时代的先锋，同时也是 DT 时代最大的玩家，其大数据应用技术及需求同样领先其他行业。数据行业最早由 Google, Facebook 等国外互联网厂商的大数据需求所驱动，而中国大数据市场仍然由互联网厂商主导，其中各方面都处于领先地位的毋庸置疑是 BAT 三家。目前最容易获取到的数据大都来自 BAT，同时一直处于信息技术前沿，凭借多年海量数据积累了大数据

资源对内对外应用实力，因此他们对于数据交换、共享、合作持非常开放的合作态度。

在大数据应用上面，百度、腾讯在 2015 年就完成了平台的开发上线，阿里也于 16 年推出了大数据平台“数加”。BAT 都是利用积累已久的海量数据和技术能力，提供开放的对外接口，助力传统行业实现大数据应用的落地和突破，目前针对的行业包括 O2O、零售、旅游、房地产、金融、保险等。海量数据积累、目标用户分析、前沿模型算法、高效计算能力；被转化为通用性的产品服务组件--行业洞察、营销决策、客群分析、舆情监控、店铺分析、推荐引擎以及数据加油站。互联网厂商的大数据平台已经能够渗透到各行各业，实现部分行业的趋势洞察、客群精准触达、精准营销决策、风险危机防控等核心价值，打开更大的市场空间，成为其新的业务增长点。

大数据的处理、分析、应用能力在与日俱增的数据量背景下以及更多个性化的数据分析需求的驱动下，将决定互联网行业未来版图。

(4) 大数据变现的“源”与“流”

从大数据产业链价值来看，最终大数据的变现机会将会趋近“源”与“流”。源就是指数据源，从行业的定义来看，数据源本身的巨大价值是一切的源头。而海量数据的拥有者将无疑最有可能变现更多价值。而流则是指数据挖掘手段，它将决定数据变现通道的宽度和广度，形成数据源导流变现的最佳渠道。当前的数据分析依然处于初级阶段，其对各个行业应用的渗透和推动并未形成显著效应。一旦数据挖掘能够更有效应对和处理结构化、半结构化以及非结构化数据，对于下游应用的作用将会呈指数级的增强，从而形成跨越式发展。

(5) 数据安全的重新定义

近年来信息安全事件频发，以数据泄漏、信息盗取为主导导致的商业损失难以估算，一次次反向证明了数据的巨大价值。在数据化时代，企业的数据越来越宝贵，数据架构和安全设备也越来越复杂，传统的安全防护能力明显不能抵御数据黑客的侵袭。大数据本身的迅速发展也助长了数据黑客的盗取能力，顺延发展出的黑色产业链已经非常成熟。传统以防御为核心的安全策略已经不再适用，唯有通过海量数据的深度挖掘与学习，才能适应千变万化的安全威胁，并实现由“被动防御”到“主动智能”的信息安全战略升级。基于大数据框架对企业的系统、应

用和用户访问行为数据进行存储与分析，并采用机器学习和算法来检测异常行为，是业界公认的抵御新型外部攻击（APT，Advanced Persistent Threat）和内部人员恶意窃取核心数据（Insider Attacks）的最有效方式，可以最大限度的保护信息资产的安全。

三、大数据行业的上下游

（1）上游行业

大数据行业的上游主要是硬件基础设施类厂商例如 ICT 以及基础能力类软件提供商如 Hadoop、Spark。目前主流的硬件和基础软件厂商仍然以欧美厂商为主，国内厂商基本处于市场和技术的第二阵营。无论是处理器、存储、云平台、物联网等各个细分领域竞争都比较激烈，产品升级非常频繁，其市场化程度也较高，整体行业格局趋于稳定。这样也导致部署大数据的基础硬件平台和工具平台软件不存在完全垄断，选择度较大，处于性价比提升的阶段。

（2）下游行业

大数据行业的下游行业覆盖面非常广泛。在大数据发展早期利用大数据进行精准营销、开店选址等应用曾是大数据变现的唯一方式。虽然，大数据在其他行业的前景都被广为普遍被看好，但是受限于技术以及基础设施的不成熟，相关应用多停留在研究阶段。2015 年随着产业生态的逐渐完善，大数据在健康医疗、征信金融、民生工程等行业都实现了应用落地。大数据与云计算是被绑定的一体两面，只有在信息化程度较高的行业中才有可能实现云平台的架构，从而达到大数据运算能力的要求。导致当前国内大数据应用的下游将主要集中在金融、能源、安全、交通和政府等信息化程度相对较高的领域。而随着国内企业信息化水平的普遍提升，以及企业对大数据接受程度的提高，大数据行业的下游覆盖面将逐步拓宽，对数据分析产品与服务的刚性需求将保持稳定增长。

四、大数据行业发展的有利和不利因素

1、有利因素

（1）国家政策支持

大数据行业作为推动技术创新、管理创新和产品创新的主要动力，17 年 1 月 17 日工业和信息化部正式印发《大数据产业发展规划(2016-2020 年)》，全面部署十三五（2016-2020）时期大数据产业发展蓝图：行业收入破万亿、行业年

均复合增速保持在 30%、发展 10 家核心龙头、500 家应用企业、建设 10-15 个大数据综合实验区等具体指标的制定，将使得整个产业从资源开放、基础设施建设以及应用领域等几个层面得到支持，提振产业发展的信心。为助力其他传统行业深度转型，实现供给侧改革的发展营造良好的数据环境。

(2) 大数据为保障经济增长，行业持续发展提供动力

根据《国民经济和社会发展第十三个五年规划纲要》，未来五年，GDP 年均增速要保持在 6.5% 以上。到 2020 年经济总量要超过 92.7 万亿元。到 2020 年国内生产总值和城乡居民人均收入比 2010 年翻一番，主要经济指标平衡协调，发展质量和效益明显提高，服务业占 GDP 比重则要从 2015 年的 50.5% 进一步提升至 56%。未来的经济增速将主要来自传统产业的转型升级和服务业的高速发展，而大数据行业则是为这一发展提供持续动力和有力保障。不仅能给传统产业提供销售升级的保证，同时也能提供新业务方向的指导，更能为服务业提供优质的技术支持和拓展空间。

(3) 各地区和产业部门信息化进程不断深入，为大数据行业创造了广阔的行业市场空间

2016 年，继国家发改委印发了《关于组织实施促进大数据发展重大工程的通知》后，环保部 3 月 7 日正式印发《生态环境大数据建设总体方案》、6 月 24 日国务院办公厅印发《关于促进和规范健康医疗大数据应用发展的指导意见》、国土资源部、国家林业局、煤工委、交通运输部、农业部也都推出了行业内大数据发展意见和方案……。各大产业、各细分领域的政策和实施方案，充分推动了大数据脱虚向实，从规划逐渐向行业落地的过程。各行各业的需求打开了大数据产业的实际应用之路，后续的市场空间非常广阔。

(4) 消费升级带来巨大市场机遇

当前的国内经济增速开始放缓，处于低速稳定发展阶段，但是国内人均可支配收入则仍保持增速不断上升。根据国家统计局数据，2016 年全年全国居民人均可支配收入 23,821 元，比上年名义增长 8.4%，扣除价格因素实际增长 6.3%（这个值的中位数为 20,883 元，比上年名义增长 8.3%）。消费升级已成定势，而供给侧的改革却缺少方向与实际的抓手，而大数据可以从消费趋势和供给结构的数据中挖掘出最有价值的参考指标，为深化消费升级和供给改革带来指向性的帮助。

当前已经成熟的大数据指导多数在精准营销和价值判断上,作为大数据行业变现的启爆点已经说明了行业的价值和能效。可以预见后续在风险控制、安全防护、身份鉴别等方向都会有成熟的应用出现。大数据不断扩张其涉入的领域,进一步深入行业内部挖掘需求,而消费升级带来的巨大市场将大数据应用全面推向规模化发展,形成巨大的产业规模。

2、不利因素

(1) 全面发展大数据仍需要时间

大数据行业的发展是一个系统工程,并非有部分的专利技术就能拉动全部产业链的。从上游的服务器等硬件,到提供平台基础服务能力的底层软件,再到上层定制化的功能应用,每一个环节都相互关联,形成一个链条,缺少了一个就不可能承受更大的拉力。这也造成了即使有需求的拉动,也不能实时产生联动效应;也是为什么大数据针对细分行业的应用必须有时间的磨合。而且由于传统行业和行业中的大中型企业的信息化程度普遍不高,数据的分散程度较高且不能顺畅的流通,导致大数据进入到其中并起作用需要一段比较长的时间。

(2) 高端设计人才储备相对缺乏

大数据行业是计算机软件、硬件、集成技术、数据清洗、数据挖掘及可视化等多个学科的智力密集型高科技行业。作为新兴行业,人才对于大数据企业在行业竞争中取得优势地位具有至关重要的作用。目前,我国对于各个细分方向上的高端专业人才的培养和输送尚不能满足行业快速发展的需求,更何况技术的不断升级进步还需要多方向研发的复合型技术人才。因此,高端设计人才的短缺将成为行业发展过程中亟待解决的重要课题。

(3) 经济下行压力短期内难以改善

2014年以来我国宏观经济形势面临较大下行压力,虽然经济短期内发生巨大风险的可能性不大,但是宏观面在短期内难以得到改善。“稳增长”政策受制于地方政府债务压力较难有爆发性增长,最艰难的供给侧改革处在攻坚阶段,大数据行业的发展也被动受限制。无论是从固定资产的投入,还是技术人才资源的流动,在下行环境中新兴产业的发展都会受到一定程度的波及。而是否能在经济压力下,利用大数据落地找到发展突破口将成为行业进入高速成长期的关键。

五、大数据行业风险特征

1、新兴行业风险

大数据行业是一个新兴行业其中的大部分企业都属于初创期，都存在自身发展核心动力不足、抗风险能力差、资源整合能力欠缺等问题，亦制约了行业的发展速度。目前国内行业龙头包括华为、百度、腾讯、阿里等在大数据各个细分领域都占据了一些份额，但是从产品整体的质量和服务能力上来说仍处于初级阶段，相比国外成熟厂商的产品还有很大的距离。但是国内大数据资源丰富，是全球最重要大数据市场之一，创造的数据规模远超其他国家，这为中国大数据产业提供了丰富资源，为国内企业发展弯道超车创造良好机会。

2、人才市场风险

数据行业的人才缺乏是一个全球性的问题，在硅谷的深度学习人才年薪 300 万已经是一个旧闻了。麦肯锡公司的预计显示，美国到 2018 年深度数据分析人才缺口将达 14 万-19 万人，能够分析数据帮助公司获得经济效益的技术及管理人才有 150 万人的缺口。但是，目前拥有数据分析技能的专业人员严重短缺，只有三分之一的新工作岗位能招聘到人员。在全球人才汇集地的人力资源竞争已经如此激烈更勿论在高端人才极度缺乏的国内了。数据人才稀缺，企业对这些人才的争夺比较激烈。而一些在内部管理、分配制度、薪酬激励等方面存在问题的设计企业，高端优秀人才流失现象日益严重；这在一定程度上制约了行业内企业的发展速度。

3、数据安全风险

大数据孕育着很多商机，如何在商业化的过程中既保证个性化，又保护隐私这是一个对立的问题，给应用开发提出很大挑战。如何防止数据被过度泛滥、被公开和被不法分子利用，这是大数据行业发展过程中必须面对的问题。近年来在全球不断出现数据泄露带来安全风险背景下，大数据行业的发展受到一定程度上的限制，对业内企业的业务造成极大的负面影响。未来国家对数据安全防护政策仍存在较大的不确定性，所以会存在目标市场行业数据被管控的风险，进而影响目标行业应用的整体运营。如果管控范围扩大或者政策规划未厘清，将对大数据行业的整体运行产生不利影响。