

# 北京市大数据产业发展研究报告

数据已成为继物质、能源之后的第三大战略资源，大数据时代的到来标志着信息化已进入了以数据的深度挖掘与融合应用为特征的智能化阶段（信息 3.0）。当前，全球信息技术产业体系面临重构，大数据技术成为信息技术的核心，掌握大数据也就掌握了信息的“控制权”。

近年来，国务院和北京市相继出台了促进大数据发展的行动纲要和计划，推动大数据发展与各行业领域的深度融合，大数据成为经济社会创新发展的重要引擎，成为稳增长、促改革、调结构、惠民生的重要手段。

## 一、大数据产业概述

大数据产业是指以数据采集、存储、管理、分析、展示、应用为主的各类经济活动，包括数据资源建设、大数据软硬件产品的开发、销售和租赁活动，以及相关信息技术服务。大数据产业主要包括五个层面：（1）数据资源，数据市场、网络数据、地理信息数据、企业数据、政府数据、医疗数据等；（2）基础架构，包括网络服务、底层技术、安全、云服务平台等；（3）数据服务，从数据处理流程可以分为数据采集和预处理、存储与管理、数据挖局与分析、数据可视化、数据安全等；从产业生态可以分为可视化、应用分析、社交分析、电商分析、商业智能（BI）、AI 技术、分析服务等；（4）融合应用，包含了与政务、工业、医疗、金融、交通和电信等行业紧密相关的应用软件和整体解决方案；（5）周边服务，包括研究教育机构、产业联盟、社区媒体、咨询机构等。



图 1 大数据产业生态系统

## 二、国际大数据产业发展现状和趋势

### (一) 发达国家大数据战略持续拓展

掌握和运用大数据的能力成为判断国家竞争力的重要依据，各发达国家纷纷将大数据作为国家发展战略。美国高度重视大数据研发和应用，2012年3月推出“大数据研究与发展倡议”，将大数据作为国家重要的战略资源进行管理和应用；2016年5月进一步发布“联邦大数据研究与开发计划”，不断加强在大数据研发和应用方面的布局；2019年6月发布了《联邦数据战略第一年度行动计划》草案，旨在编纂联邦机构如何利用计划、统计和任务支持数据作为战略资产来发展经济、提高联邦政府的效率、促进监督和提高透明度。欧盟2014年推出了“数据驱动的经济”战略，倡导欧洲各国抢抓大数据发展机遇；2018年实施《通用数据保护条例》，强化数据安全和个人信息保护；2019年初，欧洲议会通过了一项决议敦促欧盟及其成员国创造一个“繁荣的数据驱动经济”，预计到2020年欧盟国内生产总值将因更好

的数据使用而增加1.9%。此外，英国、韩国、日本等国也出台了类似政策，推动大数据应用，拉动产业发展。

## （二）大数据产业生态基本形成

全球各大信息技术巨头企业非常重视大数据发展机遇，通过自身研发、资源整合、合作、并购等途径，强化大数据领域的技术水平和应用能力，纷纷推出大数据相关产品和服务。谷歌、微软等企业在内存数据库、企业大数据分析、大数据安全等技术领域取得一系列突破，引领全球技术创新。大数据应用在发达国家政务、金融、医疗、工业、教育和中小企业等重要领域先后落地，实践案例不断丰富。大数据产业规模平稳增长，据国际权威机构Statista于2019年8月发布的报告显示，预计到2020年，全球大数据市场的收入规模将达到560亿美元，较2018年的预期水平增长约33.33%，较2016年的市场收入规模翻一倍。

## （三）大数据技术面临颠覆式发展机遇

1、大数据应用需求驱动计算技术体系重构。具体包括以下几个方面：新型存储介质及新型运算器件的涌现导致计算机体系结构的变革；云模式成为大数据处理的新趋势，云计算呈现应用领域化、资源泛在化、系统平台化的发展态势；计算能力从云向边扩散，计算向数据靠近，形成云端融合的计算新模式；数据及应用需求的多样性，使得计算平台的软件定义成为主流；多源异构数据的一体化访问、面向分析的软硬件协同等需求亟需新一代数据管理技术与系统；高维、流式、语义化的大数据分析需要新的方法和工具。

2、网络通信向宽带、移动、泛在发展，加速数据汇聚。一方面，海量大数据的高速传输需求推动网络技术向 Pb/s 发展，软件定义网

络成为趋势；另一方面，上千亿设备的网络接入需求，推动 Gb/s 级高密度泛在移动通信网络发展。

3、渐进式技术发展面临极限，亟需基础器件的突破。高性能、高时效、高吞吐等极端化大数据应用需求的出现，亟需高通量处理芯片、高密度非易失存储器件、多通道数据可视化设备等基础器件创新与变革。

4、软硬件开源开放将重构产业发展新生态。随着 IT 产品与服务从系统为中心向数据为中心转移，软硬件不再成为制约 IT 产品和服务的关键要素，软硬件开源成为创新能力提升的新模式，并由此重构产业发展的新生态。

#### （四）数据合规性要求日益严格

近两年来，各国在数据合规性方面的重视程度越来越高，但数据合规的进程仍任重道远。2019年5月25日，旨在保护欧盟公民的个人数据、对企业的数据处理提出了严格要求的《通用数据保护条例》（GDPR）实施满一周年，数据保护相关的案例与公开事件数量攀升，同时也引起了诸多争议。

GDPR的正式实施带来了全球隐私保护立法的热潮，并成功提升了社会各领域对于数据保护的重视。例如，2020年1月美国加州的消费者隐私法案（CCPA）正式生效。与GDPR类似，CCPA将对所有和美国加州居民有业务的数据商业行为进行监管。CCPA在适用监管的标准上比GDPR更宽松，但是一旦满足被监管的标准，违法企业受到的惩罚更大。2019年8月，IAPP/OneTrust对部分美国企业进行了CCPA准备度调查，结果显示，74%的受访者认为他们的企业应该遵守CCPA，但

只有大约2%的受访者认为他们的企业已经完全做好了应对CCPA的准备。除加州CCPA外，更多的法案正在美国纽约州等多个州陆续生效。

### 三、我国大数据产业发展现状

#### （一）大数据产业政策环境构建完善

在政策方面，我国从中央到地方的大数据政策体系已经基本完善，目前已经进入落地实施阶段。自从2014年“大数据”这个词写入政府工作报告以来，我国大数据发展的政策环境掀开了全新的篇章。在顶层设计上，国务院《促进大数据发展行动纲要》对政务数据共享开放、产业发展和安全三方面做了总体部署。《政务信息资源共享管理暂行办法》《大数据产业发展规划(2016-2020)》等文件也都已经出台。十九大报告中提出“推动大数据与实体经济深度融合”，“十三五”规划中提出“实施国家大数据战略”。卫健、农业、环保、检察、税务等部门还出台了领域大数据发展的具体政策。截至目前，除港澳台外全国31个省级单位均已发布了推进大数据产业发展的相关文件。可以说，大数据的政策体系已经基本搭建完成，目前已经纷纷进入落地实施甚至评估检查阶段。

#### （二）大数据技术产品水平持续提升

从产品角度来看，目前大数据技术产品主要包括大数据基础类技术产品、分析类技术产品、管理类技术产品等。我国在这些方面都取得了一定的进展。

##### 1、我国大数据基础类技术产品市场成熟度相对较高

一是供应商越来越多，从最早只有几家大型互联网公司发展到目前的几十家公司可以提供相应产品，覆盖了互联网、金融、电信、电力、铁路、石化、军工等不同行业；二是产品功能日益完善，根据中

国信通院的测试，分布式批处理平台、分布式流处理平台类的参评产品功能项通过率均在 95% 以上；三是大规模部署能力有很大突破，例如阿里云 MaxCompute 通过了 10000 节点批处理平台基础能力测试，华为 GaussDB 通过了 512 台物理节点的分析型数据库基础能力测试；四是自主研发意识不断提高，目前有很多基础类产品源自对于开源产品进行的二次开发，特别是分布式批处理平台、流处理平台等产品九成以上基于已有开源产品开发。

## 2、我国大数据分析类技术产品发展迅速，个性化与实用性趋势明显

一是满足跨行业需求的通用数据分析工具类产品逐渐应运而生，如百度的机器学习平台 Jarvis、阿里云的机器学习平台 PAI 等；二是随着深度学习技术的相应发展，数据挖掘平台从以往只支持传统机器学习算法转变为额外支持深度学习算法以及 GPU 计算加速能力；三是数据分析类产品易用性进一步提升，大部分产品都拥有直观的可视化界面以及简洁便利的交互操作方式。

## 3、我国大数据管理类技术产品处于市场形成的初期

目前，国内常见的大数据管理类软件有 20 多款。数据管理类产品虽然涉及的内容庞杂，但技术实现难度相对较低，一些开源软件如 Kettle、Sqoop 和 Nifi 等，为数据集成工具提供了开发基础。中国信通院测试结果显示，参照囊括功能全集的大数据管理软件评测标准，所有参评产品符合程度均在 90% 以下。随着数据资产的重要性日益突出，数据管理类软件的地位也将越来越重要，未来将机器学习、区块链等新技术与数据管理需求结合，还有很大的发展空间。

### （三）大数据行业应用不断深化

数年之前，大数据的应用还主要在互联网、营销、广告领域。而随着大数据工具的门槛降低以及企业数据意识的不断提升，越来越多的行业开始尝到大数据带来的“甜头”。近几年，无论是从新增企业数量、融资规模还是应用热度来说，与大数据结合紧密的行业逐步向工业、政务、电信、交通、金融、医疗、教育等领域广泛渗透，应用逐渐向生产、物流、供应链等核心业务延伸，涌现了一批大数据典型应用，企业应用大数据的能力逐渐增强。电力、铁路、石化等实体经济领域龙头企业不断完善自身大数据平台建设，持续加强数据治理，构建起以数据为核心驱动力的创新能力，行业应用“脱虚向实”趋势明显，大数据与实体经济深度融合不断加深。

#### 四、大数据产业是数字经济的基础要素

近年来，我国数字经济发展迅速，生态体系正在加速形成，而大数据已成为数字经济这种全新经济形态的关键生产要素。大数据作为一种基础性和战略性资源，是提升民众生活品质、国家治理能力的“富矿”。国家高度重视的“新基建”战略布局中，大数据中心的建立，也是希望助力牢固信息化“地基”建设，推动数据要素参与到更多价值的创造和分配，保障高质量的社会发展。

##### （一）大数据是数字经济的关键生产要素

随着信息通信技术的广泛运用，以及新模式、新业态的不断涌现，人类的社会生产生活方式正在发生深刻的变革，数字经济作为一种全新的社会经济形态，正逐渐成为全球经济增长重要的驱动力。历史证明，每一次人类社会重大的经济形态变革，必然产生新生产要素，形成先进生产力，如同农业时代以土地和劳动力、工业时代以资本为新

的生产要素一样，数字经济作为继农业经济、工业经济之后的一种新兴经济社会发展形态，也将产生新的生产要素。

数字经济与农业经济、工业经济不同，它是以新一代信息技术为基础，以海量数据的互联和应用为核心，将数据资源融入产业创新和升级各个环节的新经济形态。一方面信息技术与经济社会的交汇融合，特别是物联网产业的发展引发数据迅猛增长，大数据已成为社会基础性战略资源，蕴藏着巨大潜力和能量。另一方面数据资源与产业的交汇融合促使社会生产力发生新的飞跃，大数据成为驱动整个社会运行和经济发展的新兴生产要素，在生产过程中与劳动力、土地、资本等其他生产要素协同创造社会价值。相比其他生产要素，数据资源具有的可复制、可共享、无限增长和供给的禀赋，打破了自然资源有限供给对增长的制约，为持续增长和永续发展提供了基础与可能，成为数字经济发展的关键生产要素和重要资源。

## （二）大数据是发挥数据价值的使能因素

市场经济要求生产要素商品化，以商品形式在市场上通过交易实现流动和配置，从而形成各种生产要素市场。大数据作为数字经济的关键生产要素，构建数据要素市场是发挥市场在资源配置中的决定性作用的必要条件，是发展数字经济的必然要求。2015年《促进大数据发展行动纲要》明确提出“要引导培育大数据交易市场，开展面向应用的数据交易市场试点，探索开展大数据衍生产品交易，鼓励产业链各环节的市场主体进行数据交换和交易”，大数据发展将重点推进数据流通标准和数据交易体系建设，促进数据交易、共享、转移等环节的规范有序，为构建数据要素市场，实现数据要素的市场化和自

由流动提供了可能，成为优化数据要素配置、发挥数据要素价值的关键影响因素。

大数据资源更深层次的处理和应用仍然需要使用大数据，通过大数据分析将数据转化为可用信息，是数据作为关键生产要素实现价值创造的路径演进和必然结果。从构建要素市场、实现生产要素市场化流动到数据的清洗分析，数据要素的市场价值提升和自生价值创造无不需要大数据作为支撑，大数据成为发挥数据价值的使能因素。

### （三）大数据是驱动数字经济的核心动能

推动大数据在社会经济各领域的广泛应用，加快传统产业数字化、智能化，催生数据驱动的新兴业态，能够为我国经济转型发展提供新动力。大数据是驱动数字经济创新发展的重要抓手和核心动能。

大数据驱动传统产业向数字化和智能化方向转型升级，是数字经济推动效率提升和经济结构优化的重要抓手。大数据加速渗透和应用到社会经济的各个领域，通过与传统产业进行深度融合，提升传统产业生产效率和自主创新能力，深刻变革传统产业的生产方式和管理、营销模式，驱动传统产业实现数字化转型。电信、金融、交通等服务行业利用大数据探索客户细分、风险防控、信用评价等应用，加快业务创新和产业升级步伐。工业大数据贯穿于工业的设计、工艺、生产、管理、服务等各个环节，使工业系统具备描述、诊断、预测、决策、控制等智能化功能，推动工业走向智能化。利用大数据为农作物栽培、气候分析等农业生产决策提供有力依据，提高农业生产效率，推动农业向数据驱动的智慧生产方式转型。大数据为传统产业的创新转型、优化升级提供重要支撑，引领和驱动传统产业实现数字化转型，

推动传统经济模式向形态更高级、分工更优化、结构更合理的数字经济模式演进。

当前，我国既要预防疫情再次爆发，同时要恢复生产和扩大消费，而能将这三者有机结合起来的只有数字经济，数字经济是当前经济发展主要的推动力。“十四五”时期，我国数字经济将继续快速发展全面发力。

首先，数字化消费将向新领域延伸，继续创造重量级新消费形态。网络办公、网络会展、数字学习、数字医疗、数字文化、数字传媒，以及智能家庭居住、智能个人穿戴、智能交通出行等，都将得到较快发展。

其次，数字化将向制造业和农业等生产领域扩展，向供应链、销售链扩展，提升全链条资源配置效率。例如，数字化智能化可以同时实现个性化定制和低成本制造，客户可以实时提出需求，全程参与研发生产过程。在“双循环”的背景下，数字化网络平台能够聚合产业链上多环节多种类企业和多种生产要素，为各方提供多种类型的交互机会，提供业内所需的各种服务。在特殊时期比如此次新冠肺炎疫情期间，当原有的产业链断裂时，平台可以智能化地在供需双方之间进行匹配，迅速寻找替代或调整方案，快速补链接链，使全球分工体系的稳定性安全性大大提高。

## 五、北京发展大数据技术的优势和挑战

### （一）优势

#### 1、产业链条较健全，集聚效应显现

##### （1）大数据资源市场日趋活跃

以中关村数海大数据交易平台为代表的联盟性质交易平台，和以数据堂等公司为代表的互联网综合数据交易服务平台，极大促进了北京乃至全国的大数据交易市场的发展。此外，在地理信息、网络、贸易、房地产、人口数据领域均有企业布局。

#### （2）基础架构类企业实力雄厚

在基础架构领域，北京市骨干软硬件企业陆续推出自主研发的大数据基础平台产品。云服务企业初具规模，引领行业发展，以金山云、百度网讯、光环新网、荣之联为突出代表；存储管理类企业发展迅速，以人大金仓、博彦、平凯科技等为代表；数据安全领域神州绿盟达到10亿元规模。

#### （3）数据服务类企业引领前沿发展

北京市在数据获取、数据存储、数据分析、数据安全等产业环节涌现出一批领先企业和创业企业。领头企业积极布局深度学习等前沿技术，在语音识别、图像理解、文本挖掘等方面抢占技术制高点，如捷通华声、旷视、云知声等；更多企业在数据分析（如拓尔思、百分点）、商业智能（如用友网络）、应用分析（如七麦科技）等领域崭露头角。

#### （4）融合应用类企业发展全面

政府大数据方面，在精准扶贫、社会保障等领域，通过大数据技术和产品，普遍提升了公共资源优化配置和服务水平，代表企业有九次方、太极等。健康医疗大数据方面，依托北京大学医疗健康中心数据，北京大数据研究院与跨国制药企业阿斯利康共建健康医疗大数据实验室。金融大数据方面，积极开展精准营销和大数据风控等应用实践；文化旅游大数据方面，滴滴出行规模最大。文化传媒大数据方面，

代表性企业有爱奇艺、搜狐、新浪、今日头条等。农业大数据方面，有上市公司大北农。工业大数据方面，本地企业也在积极运用大数据服务，挖掘数据价值，如和利时、北科天绘、金控数据、树根互联等。

#### (5) 大数据支撑平台丰富多样

北京市大数据支撑平台包括联盟协会、科研教育、社区媒体、产业园区等。已初步形成科技创新驱动大数据落地的“北京模式”，“人才高地”、“技术引擎”和“孵化器”作用日益显著。以中关村大数据产业联盟为代表的中介服务机构蓬勃发展，极大促进了大数据产业的发展。

### 2、技术处于国内领先，竞争能力明显提升

北京汇聚了百度、神州数码、字节跳动和京东等掌握领先技术的企业，利用软硬件技术基础和掌握大量数据资源的优势，不断组织大数据关键技术攻关。百度成立了数据研究中心，基于搜索引擎天然优势和技术实力，开发出很多大数据技术和产品，“百度大数据引擎”是全球首个开放大数据引擎。计算所、金山云、奇虎联合研制基于国产智能芯片的大数据云平台。北京大数据研究院研发的“新一代指纹识别技术”实现了一系列技术突破，成为国际上首次实现指纹的无标注对比，以及重点技术成果“分子动力学模拟平台”是基础科学的重要研究工具，解决了分子模拟精度与效率难以兼顾的难题。北京理工大学、北京大学、太极计算共同突破了大数据共享融合关键技术的研发并进行了政务治理验证。北京轨道交通建设公等共同研发出基于大数据的智能化车辆基地管控系统关键技术。

### 3、研发基础雄厚，创新体系初步建立

北京聚集了北大、清华、北航、中科院等全国大数据骨干研究单位，大数据分析技术国家工程实验室等 7 家国家工程实验室，1 家国家级大数据国际科技合作基地，15 家北京市重点实验室，7 家北京市工程技术研究中心。建立了整合官、产、学、研多方面资源的大数据研究院，进行科学研究与人才培养。该领域的核心专家团队有北京大学鄂维南院士、北京理工大学梅宏院士、清华大学孙家广院士、北京邮电大学李德毅院士、中科院计算所李国杰院士、中科院生物物理研究所陈润生院士、北京大学高文院士、中科院李未等 8 位院士领衔的研究团队。

#### 4、大数据产业发展态势良好

北京大数据产业规模日趋壮大，全市大数据企业和机构数量达到 620 余家，主要集中在中关村科学城，其中上市公司 80 余家，独角兽企业 40 余家。从投融资方面看，2019 年北京大数据产业融资次数共 237 次，涉及企业 89 家。从投资领域分布看，IT 服务和互联网领域是最热的投资领域，其他相对热点的领域还包括金融、医疗等。

#### （二）挑战

一是原创技术仍不多见。由于缺乏原创性大数据技术和产品，大多企业只能借鉴和引用国内外现有模式，前瞻性技术研发方面仍处于跟随状态，导致企业间相似性较高，重复性竞争较大，最终使得企业规模较小、效益较差。

二是产业生态不够健全。企业、科研院所、高校之间尚未建立起知识创新、技术研发和成果转化紧密结合的有效机制。高校院所主要以课题方式获得项目资助，体量不大，且缺乏持续性，而企业投入前沿基础研究动力不足。

三是数据资源有待盘活。由于政府各部门单位的数据尚未完全对外公开，非上市企业内部数据尚未转化为资源，因此数据资源类大数据企业数量偏少，数据交易量仍不成规模。

四是缺少数据管理规范。多个政务部门、行业企业之间数据标准不统一，数据接口错综复杂，导致公共数据共享开放仍未深入展开，涉及公共服务、城市管理信息的共享，更是难以实现。

## 六、北京市发展大数据产业的工作建议

### （一）布局大数据前沿基础研究

围绕大数据分析与人机交互领域进行前瞻布局，突破数据智能的关键技术，提升大规模数据处理、机器学习、语义理解、增强分析、可视化等多技术融合能力，研制多源异构数据一体化访问、人机协同交互平台，实现智能交互界面工具。

### （二）建立重点项目库，持续研发资金投入

将大数据产业支持政策列入政府统筹重大科技资金和产业资金的支持范围，对大数据产业的重点项目实行政府全程跟踪服务，定期进行调度协调，构建稳定支持机制，择优给予重点支持。

### （三）以应用促进共性技术研发布局和创新突破

以行业应用为牵引，优先选择政府服务与监管、智能制造、医疗、交通、环保、金融等行业作为重点应用领域，引导和支持大数据企业及用户单位利用大数据提供创新应用服务，推动产业转型升级与提质增效，推进大数据技术创新、应用模式创新和商业模式创新。

### （四）把握融合发展机遇，促进产业发展

随着大数据和 5G、人工智能基础技术及应用的发展，结合多源异构大规模数据处理、数据挖掘、机器学习、人机交互、可视化等多

种技术融合的数据智能(Data Intelligence)将成为引领未来5年发展的核心技术。作为一个跨学科的研究领域和融合技术，数据智能技术重点研究预测性分析和指导性分析的关键技术，从数据中提炼、发掘、获取有揭示性和可操作性的信息，为人们在基于数据制定决策时提供有效的智能支持。而与区块链技术的融合发展可以在一定程度上解决数据确权难、数据孤岛严重、数据垄断等“先天病”。