

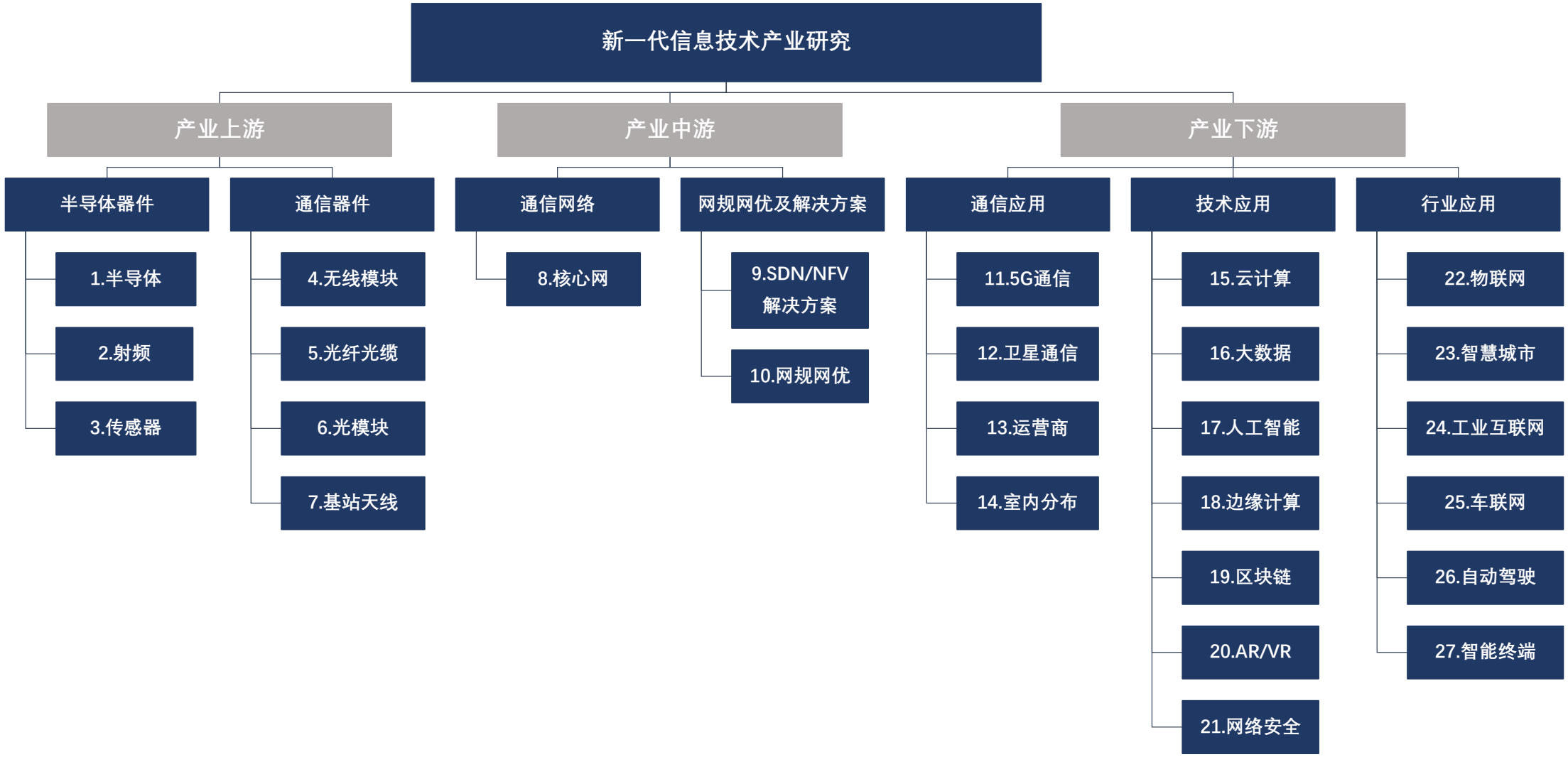
華辰資本

CELESTIAL CAPITAL

专注中国产业结构升级与创新，聚焦新一代信息技术产业发展。

2018年，在中国经济周期、产业周期、资本周期与政治周期四重叠加的特殊时期，本着“深耕产业、协同发展、价值驱动、重度赋能”的愿景，华辰资本（“华辰”）应运而生，致力成为中国最专业的创新型投资机构。

华辰资本总部位于中国最具发展活力与科技创新的深圳，专注于包括云计算、大数据、人工智能、边缘计算、工业互联网、5G等新一代信息技术领域，通过扎实的体系化产业研究与理解能力，以产业研究、投资银行、战略咨询、产业基金等模式，为新一代信息技术企业提供企业融资、战略视野、市场协同，价值管理、供应链管理、资源整合等产业赋能。



©2019 Celestial Research 华辰产业研究院

目录

一、产业分析 04

- 基本概况
- 发展历程
- 产业结构
- 应用场景
- 演进趋势
- 数据安全

二、市场分析 14

- 市场规模
- 竞争格局

三、企业分析 18

- Splunk
- Palantir
- Informatica

©2019 Celestial Research 华辰产业研究院

一、产业分析



图1 大数据的4V属性

特征	描述
数据量大(Volume)	2018年全球新产生的数据量为33ZB，中国产生7.6ZB，美国产生6.9ZB，超过人类有史以来所有印刷材料数据总量
数据类别多(Variety)	结构化数据、半结构化数据、非机构化数据
数据价值密度低(Value)	价值需要深度挖掘，原数据本身价值低
数据时效性强(Velocity)	大数据往往以数据流的形式动态、快速地产生，具有很强地时效性

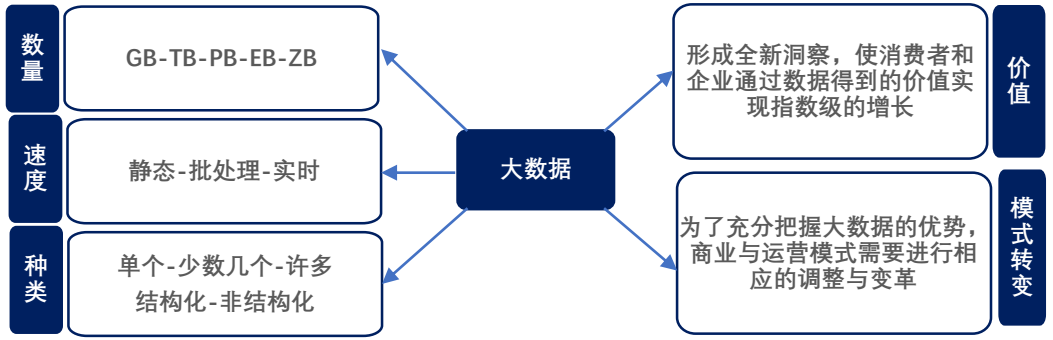
图2 大数据涉及的数据类型

数据类型	表现形式	典型场景
结构化数据	数据库表等	企业ERP、财务、HR数据库等
半结构化数据	邮件、HTML、报表等	邮件系统、网页信息、报表系统等
非结构化数据	视频、音频等	在线视频内容、音频内容、图形图像信等

图3 大数据&传统数据区别

类别	传统数据	大数据
数据对象	有限的采样样本	所有可用的数据，全数据样本
分析要求	追求结果的精确性	允许不精确和不完美，接受模糊的结论
分析结论	强调结论背后的因果关系	注重结论背后的关联关系，总结相关规则，并不关心因果关系

图4 大数据特征



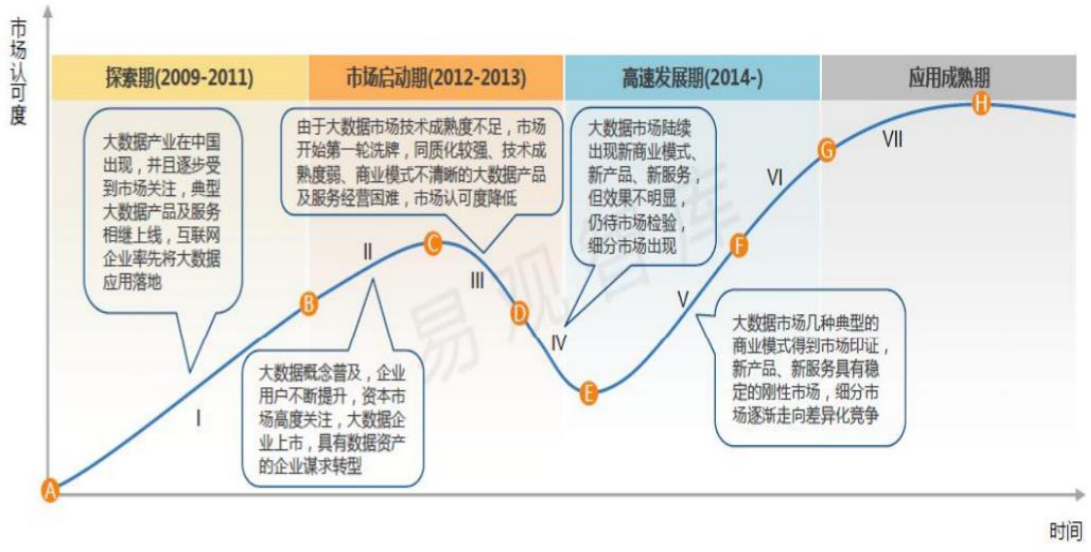
基本概况

1. **定义：**大数据指无法在一定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，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。
2. **企业家观点：**
   a. “数据将取代石油，成为未来制造业最大的能源。”——马云
   b. “云计算和大数据是一个硬币的两面，云计算是大数据的IT基础，而大数据是云计算的一个杀手级应用。” —— 张亚勤
3. **变现链条：**数据→信息→知识→决策。

资料来源：中信证券、麦肯锡、华辰资本整理

©2019 Celestial Research 华辰产业研究院

图5 中国大数据发展历程



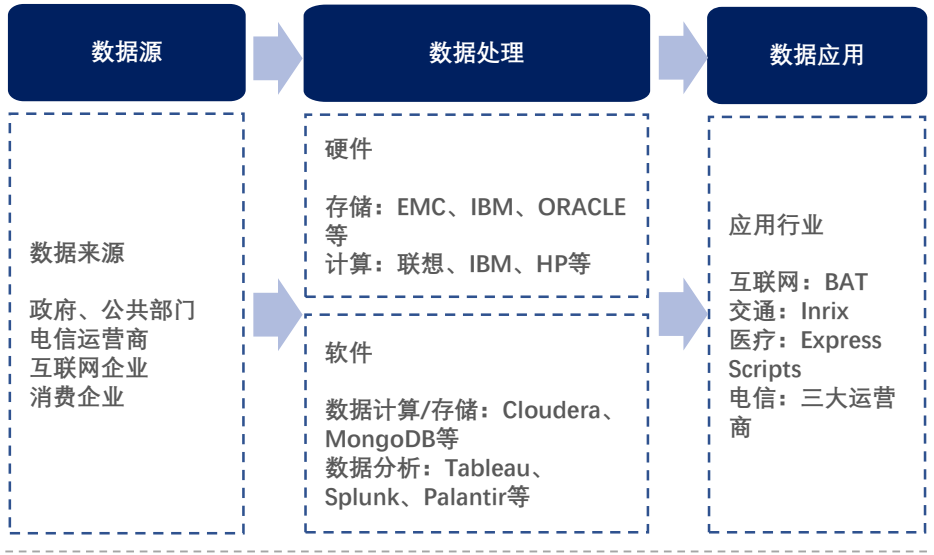
资料来源：易观智库、华辰资本整理

发展历程

1. **探索期（2009-2011）：**大数据产业在中国出现，并且逐步受到市场关注，典型大数据产品及服务相继上线，**互联网企业率先将大数据应用落地。**
2. **市场启动期（2012-2013）：**
   a. **初期，**大数据概念普及，企业用户不断提升，资本市场高度关注，大数据企业上市，**具有数据资产的企业谋求转型**；
   b. **中期，**由于大数据市场技术成熟度不足，市场**开始第一轮洗牌**，同质化较强、技术成熟度弱、商业模式不清晰的大数据产品及服务经营困难，市场认可度低；
   c. **后期，**大数据市场持续**出现新商业模式**、新产品、新服务，但效果不明显，仍待市场检验，**细分市场开始出现。**
3. **高速发展期（2014-至今）：**大数据市场几种典型的商业模式得到市场印证，新产品、新服务具有稳定的刚性市场，**细分市场逐渐走向差异化竞争**。

©2019 Celestial Research 华辰产业研究院

图6 大数据产业链



资料来源：中信证券、华辰资本整理

产业结构（从数据源到数据应用的完整生态链）

1. **上下游组成：**
   a. **上游：**数据来源提供商、存储、计算和分析涉及的硬件厂商及软件厂商；
   b. **下游：**垂直应用行业、互联网、运营商等。
2. **产业特点：**
   a. **数据源：**数据类型包括结构化、半结构化、非结构化数据；数据来源包括销售、营销等企业、从事数据整合的第三方机构和组织；
   b. **数据处理：**主要包括数据的ETL（抽取、转化、加载）、存储、计算、分析等过程，涉及数据的**切割、融合**等动作，大量中小企业通过推出特定的插件模块、基础工具等不断推动整个技术生态的繁荣；
   c. **数据应用：**该领域既包括传统的IT信息企业、统计分析企业，也包括新兴的大数据公司，比如Palantir、Inrix等，这些企业针对不同的行业、应用场景纷纷推出相应的大数据解决方案。
3. **产业应用：**电力、交通、环保、政府、金融、银行、电信、医疗、IT等行业。

©2019 Celestial Research 华辰产业研究院

图7 全球大数据产业收入结构（单位：亿美元）

类别	2011	2012	2013	2014	2015	2016E	2017E
基础硬件							
计算	1.89	2.91	4.84	6.74	8.92	10.15	11.25
存储	1.1	1.75	3.09	4.2	5.5	6.4	6.95
网络通信	0.15	0.23	0.42	0.65	0.85	1.01	1.15
小计	3.14	4.89	8.35	11.59	15.27	17.56	19.35
基础软件							
Hadoop	0.14	0.44	0.83	1.08	1.25	1.6	1.9
SQL	0.62	0.88	1.31	1.75	2.25	2.45	2.7
NoSQL	0.07	0.13	0.29	0.5	0.8	1	1.2
小计	0.83	1.45	2.43	3.33	4.3	5.05	5.8
专业服务							
专业服务	3.32	5.41	7.84	13.55	18.79	22.65	24.95
总计							
总计	7.29	11.75	18.62	28.47	38.36	45.26	50.1

资料来源：wikibon、华辰资本整理

产业模块（专业服务模块表现优异）

1. **基础硬件：**计算、存储两个子模块显示出强劲的增长动力，且整体份额占比极高，显示出云计算和大数据两个子产业较强的相互拉动作用。

2. **基础软件：**主要由Hadoop计算框架和数据管理两个子模块构成，整体规模非常有限，且增长平缓，整个数据管理模块的收入包括传统的数据软件产品收入（ORACLE、DB2等）和NoSQL商业化产品收入。

3. **专业服务：为增长最快速、份额占比最高的子模块**，整体持续保持高速增长。

©2019 Celestial Research 华辰产业研究院

图8 国内大数据应用



图9 大数据使用率每提升10%对各行业效益的提升

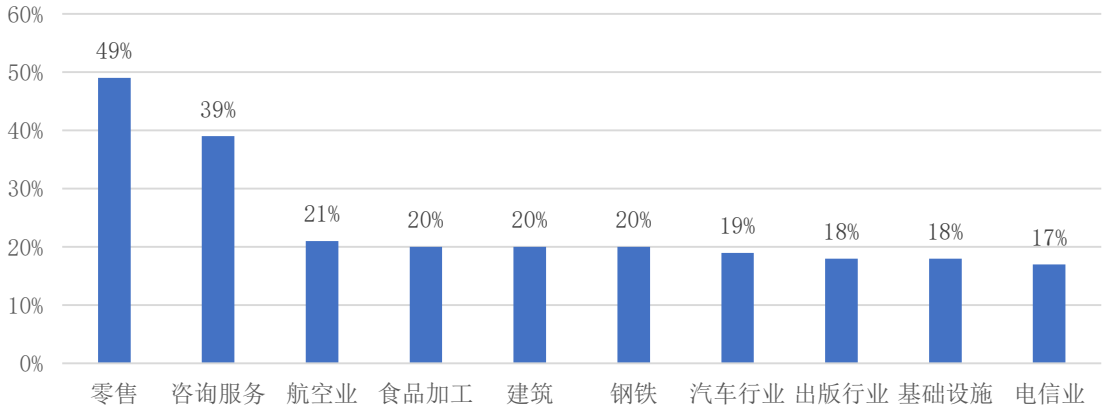


图10 大数据行业应用对应的典型应用场景

应用类别	典型应用场景
电力行业	预测电网负载扰动的可能性和发生的地点
交通行业	预测拥堵和事故在特定时间和地点可能发生的概率
环保行业	根据不同地点的废物排放，预测将来发生生物化学反应产生污染的可能性
银行业	预防客户流失、防范信用卡违约风险
电信业	预防客户流失、快速占领市场、欠费预测
公检法	网络信息挖掘、舆情监控与预测、网民个性分析

应用场景

1. 大数据的典型行业应用包括：**电力、交通、环保、银行、电信等；**
2. **互联网厂商是大数据时代的先锋，也是DT时代最大的玩家**，其中，各方面都处于领先地位的毋庸质疑是BAT三家；
3. 传统行业加速拥抱大数据，其中，**房产、汽车、医疗等**应用首当其冲；
4. 大数据使用率每提升10%，对零售、咨询服务和航空业的效益提升最大。

资料来源：易观数据、《Measuring the Business Impacts of Effective Data》美国德克萨斯大学、艾瑞咨询、华辰资本整理

©2019 Celestial Research 华辰产业研究院

图11 数据源发展趋势

类别	发展趋势
数据类型	结构化数据→视频、音频等非结构化数据共存的多类别数据
数据场景	线上数据→线上、线下数据
数据来源	企业内部数据→企业、商业、政府等第三方数据
数据关联	数据孤岛→数据连接、融合

图12 美国市场大数据来源分布

类别	数据说明
企业内部数据	结构化数据：SAP，Oracle, SQL, SQL Server, DB2, mySQL等
	非结构化数据：Cassandra, Riak, Spark, CouchDB, MongoDB等
商业数据	Marketto：（数字营销），Mailchimp（电子邮件），LinkedIn（社交&HR），Seton Healthcare（医疗），Gracenote（音乐），Optimizely（网站A/B测试），Zendesk（顾客满意度），SugarCRM（销售信息），Xero（财务信息），Factual（位置信息等）
第三方数据	社交媒体数据：Radian6, DatSift
	网络爬虫：Import.io, Kimono
	政府公开数据：Data.gov

数据源演进趋势：连接+融合

1. **大数据价值最大化需要数据的连接、融合：**
   a. **在数据类型上，**更强调结构化、半结构化、非结构化等多类别数据的采集和存储，例如音视频等；
   b. **在数据场景上**，同时注重线上、线下数据的可获得性；
   c. **在数据来源上**，除了依赖于资深的企业内部数据，更需要不断融合第三方数据；
   d. **在数据关联性上**，要消除数据孤岛，实现数据在可控范围内的自用流动和融合；
2. **单纯数据源企业出现：**当前阶段数据源环节仍和数据分析等环节紧密绑定，但未来从事单纯数据源业务的企业可能逐步出现。

资料来源：IDC、华辰资本整理

©2019 Celestial Research 华辰产业研究院

图13 传统数据分析流程

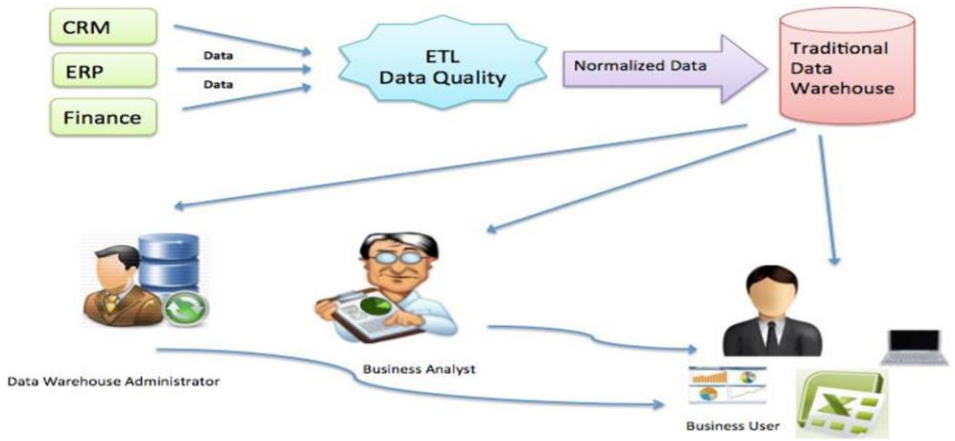
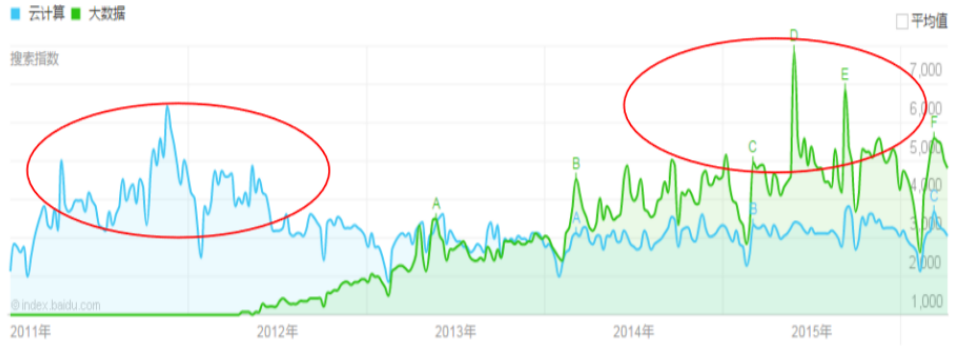


图14 云计算成熟为大数据奠定基础



数据分析演进趋势：技术变革

1. **传统数据分析时代：**通过ORACLE、DB2等数据库**集中存储形成结构化数据**，再通过BI工具和借助传统的数据挖掘算法进行分析和展现；
2. **大数据时代：**
   a. 数据量大、结构化数据占比低，传统的IT架构及分析算法难以满足需求；
   b. 2003-2004年间，Google提出**GFS**(Google File System)和**MapReduce**，奠定了**分布式文件存储系统及分布式计算模型的理论基础；**
   c. 2006年提出分布式计算和存储框架**Hadoop项目**并不断发展生态，云计算的成熟也为大数据奠定了基础。

资料来源：IDC、百度指数、Haddoop官网、华辰资本整理

©2019 Celestial Research 华辰产业研究院

图15 Tableau可视化界面



图16 Splunk业务分布



数据分析演进趋势：分析方法

1. **普适化分析：**即提供一个门槛较低的**通用软件**，例如，美股大数据龙头Tableau 开发的桌面系统，它没有强迫用户编写自定义代码，新的控制台也可完全自定义配置，从而为各行业提供简单的行业数据模板及网络共享平台；
2. **结合行业分析：向咨询、管理工具的方向发展**，利用自身多年在各行业里面积累的深刻经验并配合自身的大数据实力，在已有的数据中帮助企业挖掘更多的价值，例如，分析工具领域最成功的的新兴企业Splunk 。Splunk公司开发的基于机器数据的搜索引擎，可收集所有应用程序、服务器和移动设备（包括物理、虚拟和云端）产生的数据并生成索引，实现从一个位置快速搜索并分析所有实时数据和历史数据；
3. **未来分析方式：**未来大数据应用的集中爆发将导致数据分析需求的多样化和定制化，因此，**结合具体应用场景深入行业分析的大数据分析工具相较于通用分析工具在未来将更具有优势**。

资料来源：Tableau网站、Splunk官网、华辰资本整理

©2019 Celestial Research 华辰产业研究院

图17 大数据安全主要涉及的模块

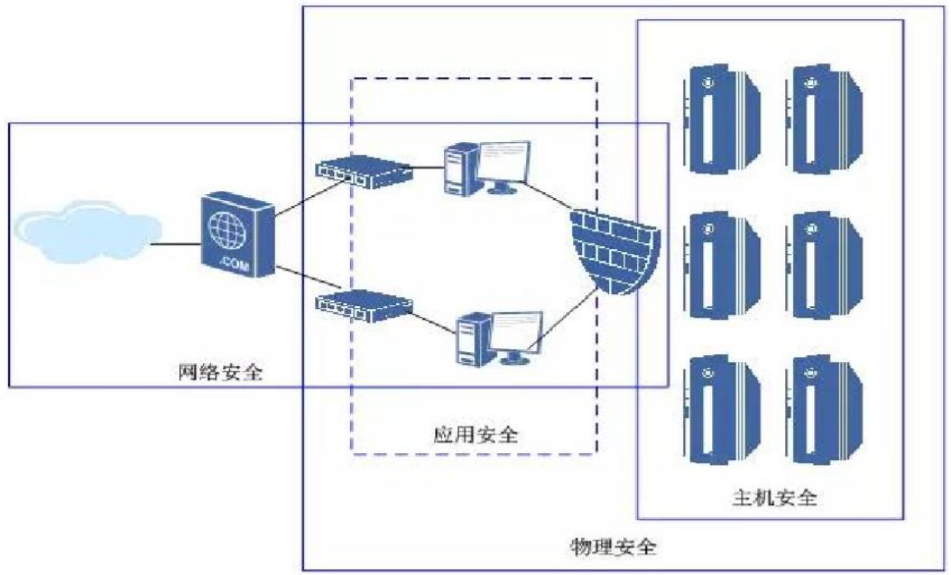


图18 大数据涉及的模块及其安全内容

模块	涉及安全内容
主机安全	主机系统、防病毒保护
网络安全	防火墙、安全通道、KERBEROS认证
应用安全	身份认证、访问控制、数据加密、数据恢复、数据容灾、日志审计、数据监控

数据安全

1. **大数据安全分析技术：**信息安全问题已变成一个大数据分析问题，大规模的安全数据需要有效地关联、分析和挖掘；未来将**出现专门的安全分析平台**，及“安全分析师”或“安全数据分析师”等岗位；
2. **隐私保护技术：**需**技术+立法**双管齐下。消费者的个人购买偏好、健康和财务情况等海量数据被收集，人们对隐私的担忧也在增大，美国奥巴马辩解说：“你不能在拥有100%安全的情况下，同时拥有100%隐私和100%便利”。现IBM、微软、柯达、花旗、Facebook、宝洁、360等企业都已经设有“首席隐私官”职位来保护用户隐私；
3. **数据脱敏技术：**在**保存数据原始特征的同时改变它的数值**，从而保护敏感数据免于未经授权的访问，同时又可以进行相关的数据处理的一种技术。借助数据脱敏，信息依旧可以被使用并与业务相关联，不会违反相关规定，而且也避免了数据泄露的风险。

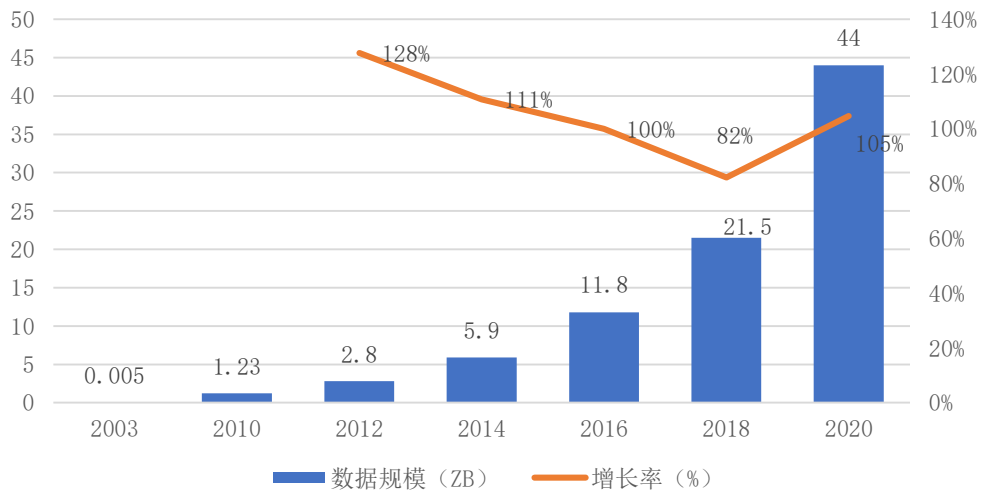
资料来源：ZTE 、华辰资本整理

©2019 Celestial Research 华辰产业研究院

二、市场分析



图19 全球产生数据规模（ZB）



资料来源：IDC、CNNIC、华辰资本整理

市场规模：数据规模爆发式增长

- 在互联网快速普及、物联网加速渗透的背景下，PC、手机、传感设备等全面兴起，推动全球数据呈现倍数增长、海量集聚的特点，为大数据产业发展奠定了庞大的数据基础：

  a. **数据规模巨大，增长极快。**根据IDC统计，2010年全球创建和复制的数据总量为**1.23ZB，**2016年这一规模为**11.8ZB，**预计2020年将达到**44ZB，**在其《数据时代2025》白皮书中，更是预测到2025年，全球创建和复制的数据总量将扩展至**163ZB**（1ZB等于1万亿GB）；

  b. **互联网与IoT带来巨大数据基础。**互联网创造了大量的线上数据，同时互联网和传统行业的融合，以及IoT时代的到来，数据的可获得途径正在被持续拓宽，互联网加速了数据间的流动和融合。

©2019 Celestial Research 华辰产业研究院

图20 全球大数据市场规模 （亿美元）

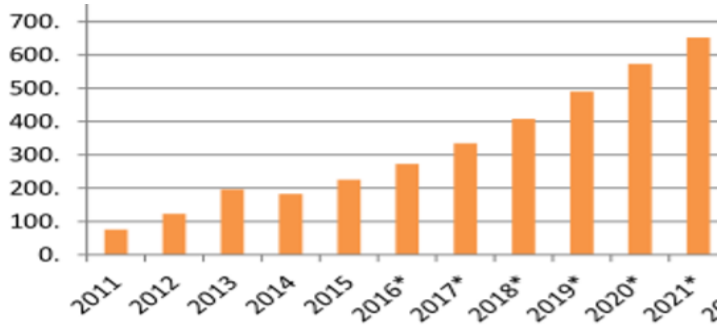
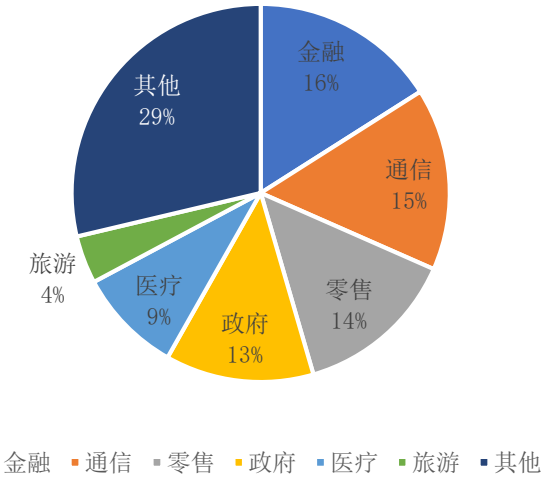


图21 中国大数据行业应用收入规模占比（%）



市场规模

1. 全球市场规模

a. 近几年，美国、欧盟、日本等主要发达经济体都积极推进各自的大数据战略。2009年，美国科学家委员会（NSTC）就发布了《开发数字数据的威力》报告;

b. 参考《大数据白皮书》、IDC、Wikibon等咨询机构分析，2016年全球大数据**核心产业规模约为300亿美元**，预计2020年有望达到近**600亿美元**。

2. 中国市场规模

a. 根据信通院数据，2017年中国大数据产业规模（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）为**4700亿元人民币**，同比增长30%，且预计2020年这一规模有望赶超**1万亿**，年均复合增速近30%;

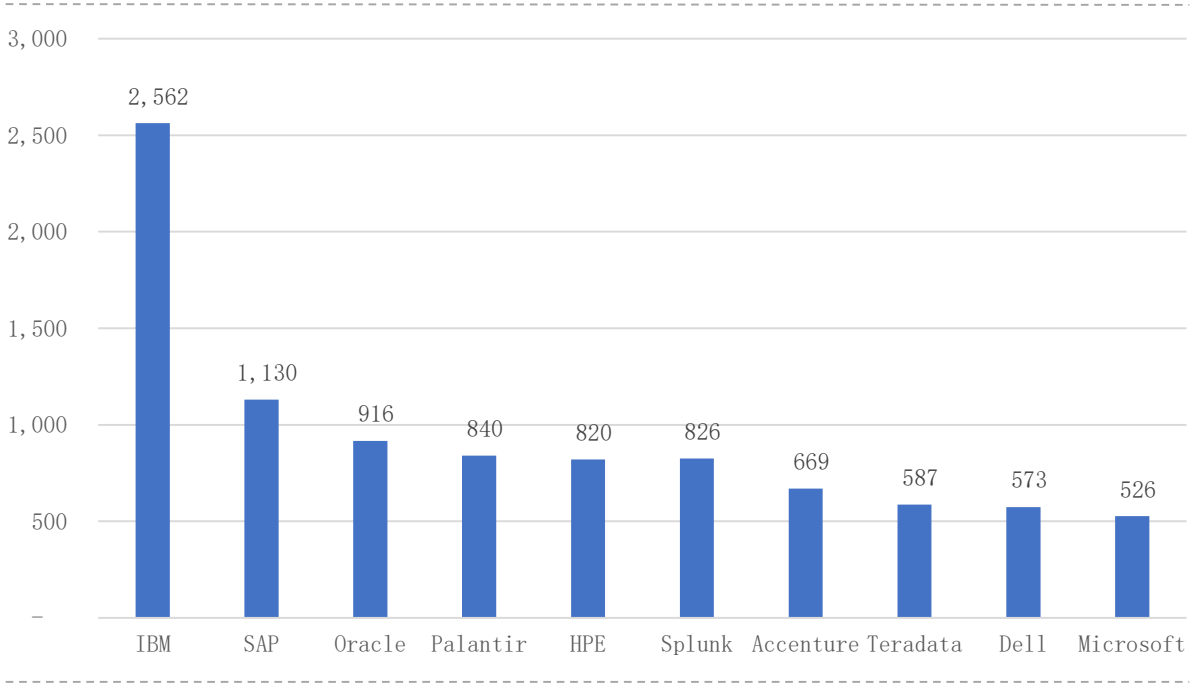
b. 大数据核心产业规模（大数据直接相关的软件、硬件及专业服务）2017年为234亿元，同比增长39%，预计2018年为329亿;

c. 从行业来看，中国大数据应用在金融行业占比最大达到16%，其次是通信行业15.6%，零售行业占比13.9%，政府占比12.7%。

资料来源：《大数据白皮书（2016）》、《大数据白皮书（2018）》、信通院、华辰资本整理

©2019 Celestial Research 华辰产业研究院

图22 2016年全球大数据竞争格局（单位：百万美元）



资料来源：前瞻产业研究源、华辰资本整理

竞争格局

- 开源技术逐步降低了企业进军大数据领域的门槛，加之数据量的规模化增长和应用场景的越发丰富，越来越多的企业加入大数据掘金浪潮，创业企业不断涌现，**互联网巨头和传统IT厂商加速投资并购以争夺市场领袖地位**；
- 数据显示，全球大数据主要由IBM、SAP、Oracle、Palantir、HPE、Splunk、Accenture、Teradata、Dell、Microsoft等占据主要市场，2016年前十大厂商份额合计占到34%，预计2020年将达到40%以上。

©2019 Celestial Research 华辰产业研究院

三、企业分析



图23 Splunk商业模式



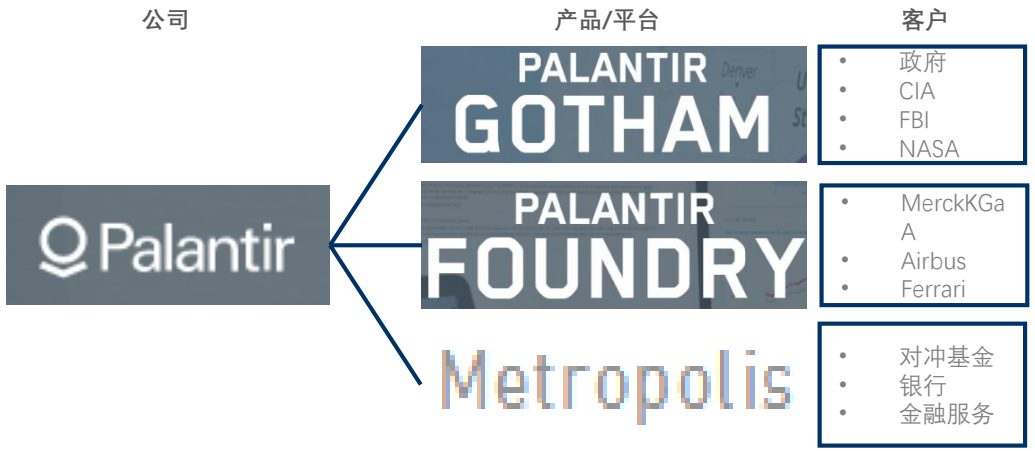
资料来源：splunk官网及公司财报、华辰资本整理

Splunk（机器数据收集分析服务商）

1. **公司概述：**成立于2003年，从机器日志分析产品起步，经历近10年的发展，逐渐成为**机器日志处理、分析领域**的全球领先企业，并通过持续的并购和外部合作，不断拓展自身技术，并围绕机器日志分析丰富公司产品组合；
2. **主要产品：**IT运营、应用管理、安全合规、网络智能与商业分析等；
3. **业务模式：软件license+运维服务。**Splunk软件平台可以实时对任何APP、服务器或网络设备的机器数据进行索引、监控与分析，并将结果生成图形化报表，并以此基础帮助客户避免服务性能降低或中断，逐渐形成软件许可证售卖和运维服务两种主要模式；
4. **财务数据：**2018年营收18亿美元，净利润-2.7亿美元，产品毛利率80%，市值170亿美元。

©2019 Celestial Research 华辰产业研究院

图24 Palantir主要产品与客户类别



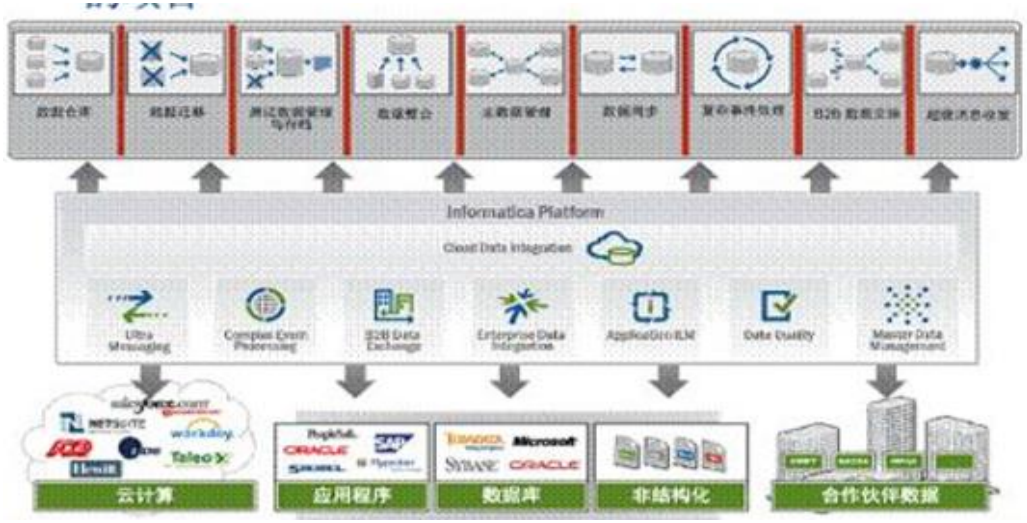
资料来源：Palantir官网、华辰资本整理

Palantir（数据聚合分析服务商）

1. **公司概述：**成立于2003年，由彼得蒂尔创立，发展初期主要面向政府机构提供数据分析服务，后借助政府订单背书，逐步将业务扩展至民用领域；2010年摩根大通成为其首批非政府客户，2011年帮助美军击毙本拉登。
2. **主要产品：**主要产品有Gotham（政府）、Foundry（企业）、Metropolis（金融），都具有开放性、可延展性和可规模性，可以数周甚至数天内完成部署、发挥作用，可以快速迭代以反映用户反馈。
3. **业务模式：**整合海量的数据源形成全方位的可视化输出，运用公司设计的相应算法进行联想、定位、分析的形式来分析单一个体或特定场景，提供解决方案。
4. **财务数据：**预计2018年营收7.5亿美元，估值410亿美元，目前有70%以上业务来源于非政府客户。

©2019 Celestial Research 华辰产业研究院

图25 Informatica 数据治理平台



资料来源：Informatica官网、华辰资本整理

Informatica（智能数据集成治理平台提供商）

1. **公司概况：**1993年创立于美国加利福尼亚州，并于1999年4月在纳斯达克上市，帮助世界各地组织为其首要业务提供及时、相关和可信的数据，从而在当今全球信息经济中获得竞争优势，2005年进入中国市场。
2. **主要产品：**最初产品数据集成软件，包括Power Center和Power Exchange，目前已经建立起了从智能数据平台到数据管理解决方案，从数据集成到数据质量管理、大数据管理、云数据管理等的完整的数据管理框架和产品线。
3. **财务数据：**2015年，微软、Salesforce等公司联合对Informatica进行私有化。2015年营业收入11亿美元，净利润1.1亿美元，估值50亿美元。

©2019 Celestial Research 华辰产业研究院

总 结

研究总结

1. 数据来源丰富多样，数据规模增长显著，大数据行业仍处于高速发展阶段，未来差异化竞争将越发突出

2. 互联网厂商掌握C端数据（消费互联），传统行业掌握B端数据（产业互联），相互融合互生将是大势所趋

3. 数据应用是数据产业链中成长空间最大的细分行业

4. 数据安全保护问题将是未来行业发展的重中之重

投资建议

1. 关注企业内部数据化建设核心团队，比如数据治理、数据中台建设到智能决策平台，以及从中衍生而来的多种SAAS工具

2. 关注企业间数据源多维度互联互通，比如工业的采集、智能运维、交通的智能调度管理、政务舆情、金融风控等

©2019 Celestial Research 华辰产业研究院

華辰資本

CELESTIAL CAPITAL

专注中国产业结构升级与创新，聚焦新一代信息技术产业发展。

联系人：欧凯

电话/微信：13510900553

邮箱：kai.ou@celestialcapital.com.cn

网址：www.celestialcapital.cn

©2019華辰資本
版权所有。

本刊物所载资料以概要方式呈现，旨在用做一般性指引，不能替代详细研究或做出专业判断。华辰资本概不对任何人士根据本刊物的任何资料采取或不采取行动而引致的损失承担任何责任。阅下应向顾问查询任何具体事宜。