

中国数据驱动型互联网企业大数据产品研究报告

2016年

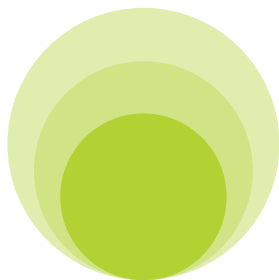
www.iresearch.cn



海量行研报告免费读

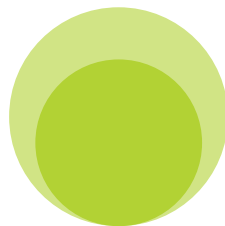
内容综述

聚焦网络媒体大数据，研究其体系架构及相关产品



大数据 Data

- **发展阶段**：从信息驱动向数据智能驱动进阶；
- **产业结构**：由数据源、技术服务及应用类型共同构成；
- **数据来源**：企业自有数据是主流数据源，数据的流通和共享将成趋势；
- **数据管理**：数据挖掘与可视化成行业热点，数据管理向垂直行业分化；
- **分析处理**：文本、图像、语音分析相对成熟，视频分析存在探索空间；
- **应用类型**：从独立数据产品向完整行业解决方案发展，定制化将成趋势。



互联网大数据 Internet

- **优势体现**：互联网基因有利于大数据的获取与处理；
- **现状分析**：互联网行业是大数据应用的领跑者。1) 已上市互联网企业中，以BAT为首的数据驱动型媒体整体表现较好；2) 未上市互联网企业中，数据驱动型企业成为领跑者；
- **机遇**：1) 手机网民持续增长，可采集群体更加广泛；2) 移动端使用时长占比超77%，信息采集更加丰富连贯；3) 场景化特征为大数据发展带来新契机；
- **挑战**：屏幕变小、行为碎片化对大数据的应用提出更高要求。



网络媒体大数据 Media

- **海外发展**：1) 企业技术探索先行，政策推动产业发展进程；2) 将自主研发与投资收购相结合，技术推动企业发展；
- **优势体现**：具有强媒体属性的网络服务的发展策略相对成熟；
- **现状分析**：中国网络媒体的数据生态系统雏形已成，多维度输出大数据能力；
- **发展策略**：大媒体以流量+广告服务为起点，推进产业生态形成。

本报告将聚焦于中国网络媒体，研究其大数据体系架构及相关产品，探讨大数据发展趋势。

中国大数据产业现状分析

1

中国网络媒体大数据产业分析

2

中国网络媒体大数据企业分析

3

中国网络媒体大数据的机遇与挑战

4

大数据定义与特征

4V：数据规模大，数据流转快，数据类型多，价值密度低

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。为了适应数据量的爆发式增长，数据处理方式正逐步向全量性、混沌性及相关性发展。为了避免信息损失，越来越多的场景倾向于对全量数据进行挖掘与分析，其中，只有一小部分数据是结构化的，而非结构化数据中同样蕴含着大量信息，因此，能够处理各式各样的混杂数据的大数据技术得到了广泛应用和迅速发展。其中，占据着重要地位的大数据预测分析，也从主要基于因果关系进行分析发展成了建立在相关关系基础上的分析模式。



Volume 数据规模大

IDC预计，到2020年，全球新建和复制的信息量将达到44ZB，中国数据量将超过8ZB。



Velocity 数据流转快

1秒定律：要在秒级时间范围内给出分析结果，超出这个时间，数据就失去价值了。



Variety 数据类型多

除了以文本为主的结构化数据、以网页数据为代表的半结构数据，也存在大量网络日志、音频、视频、图片、地理位置信息等非结构化数据。



Value 价值密度低

海量数据中，如何通过强大的机器算法，更迅速有效地完成数据的价值“提纯”，已成为目前大数据背景下亟待解决的难题。

抽样数据 >> 全量数据

精确数据 >> 混杂数据

因果关系 >> 相关关系

大数据集成与存储

数据仓库基于企业整体建立，是大数据集成和存储的基础

作为大数据集成与挖掘的基础，**数据仓库是在企业范围内共享准确一致的集成数据、针对分析需求进行数据重组的工具，基于整个企业的数据模型建立。而数据集市基于单个部门或业务主题建立，为部门业务运行提供数据服务，可以根据仓库建设也可以独立建设，形成从属数据集市和独立数据集市。**所有的从属数据集市都从属于同一个数据仓库，各子系统的数据均能保持一致。而每个独立数据集市都从各源应用环境中单独提取数据，独立拥有硬件平台、软件平台、数据和应用程序，彼此没有任何约束，无法保证数据的一致性，从长远来看，既不稳定也不可行。

数据仓库能够对海量数据进行集成，满足大数据发展对数据存储的需求，为企业提供局部解决方案和全面解决方案。

数据仓库与数据集市工作原理示意



大数据处理的数据价值流转过程



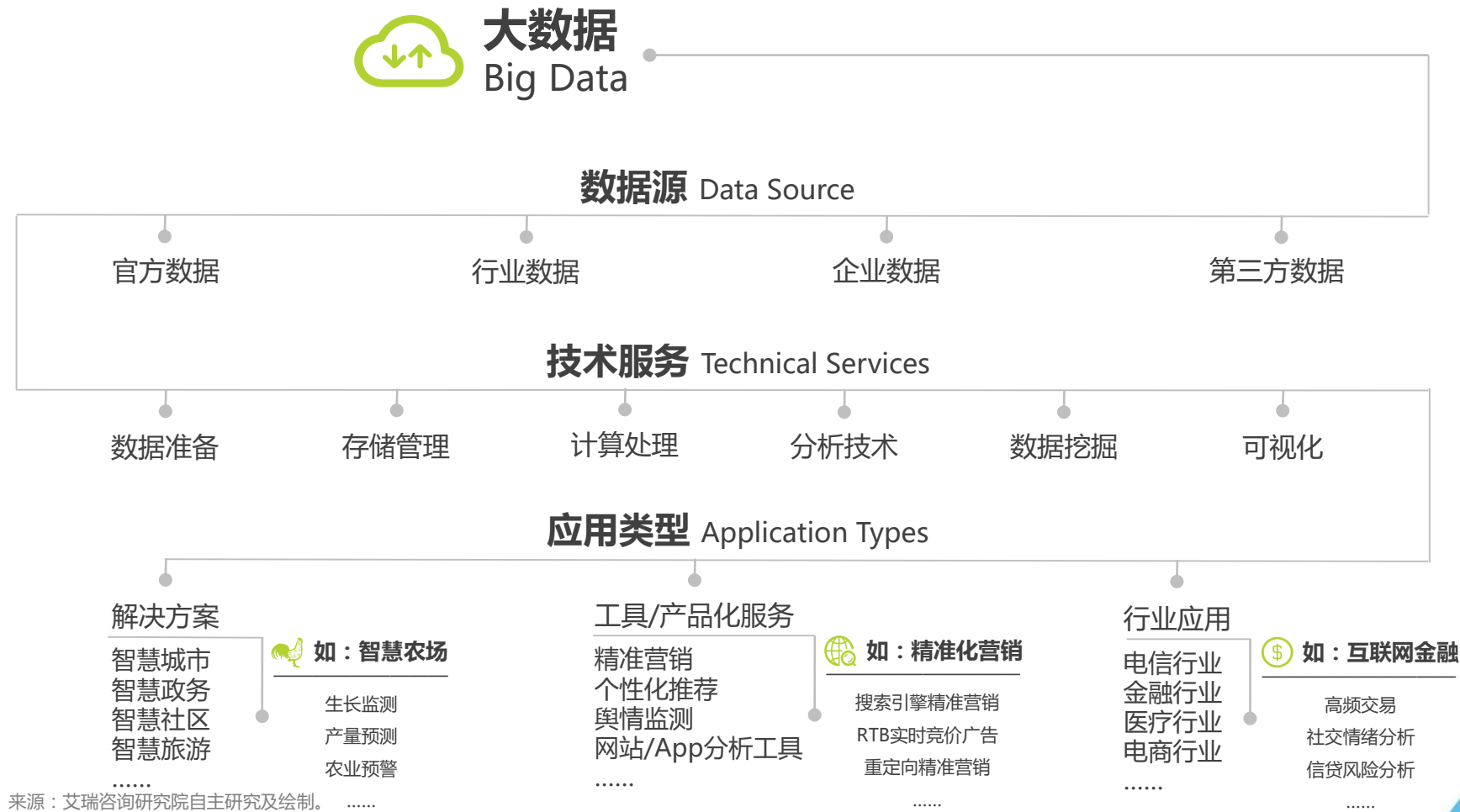
注释：ETL是指从源系统中提取数据，清洗、转换数据为一个标准的格式，并加载数据到目标数据存储区（通常是数据仓库）的工具。

来源：艾瑞咨询研究院自主研究及绘制。

大数据产业生态

产业结构由数据源、技术服务及应用类型共同构成

大数据产业结构示意图



大数据产业图谱

iResearch

艾瑞咨询



注释：此大数据产业链地图为示意图，未将所有企业标出，举例企业顺序不涉及排名，按企业中文名称首字母排列。如有分类不当，请联系：aixin@iresearch.com.cn。

大数据产业图谱

大数据应用广泛，价值升级来源于数据流通及共享

概述：本报告的大数据产业图谱以大数据产品的角度出发，对产业链角色进行划分。其中，数据管理与数据分析部分包括了大数据基础架构及相关分析技术，能够实现大数据的存储、分析、挖掘、可视化等环节。而在数据源与数据应用方面，从功能和流程角度来看具有高度的关联性及一致性，涉及行业极其广泛。

数据应用

大数据应用已渗透电信、金融、医疗、人力资源、物流、天气等多个行业，从产品角度来看，除传统的工具/产品化服务、行业应用及解决方案外，大数据应用也包括商业智能等分析服务。

- **应用情况：**目前大数据应用的主要输出形式是可视化报表，未来将在功能方面进一步融合提升。
- **典型大数据行业应用：**
 - 物流：优化整体运输流程，实现实时查询把控。
 - 营销：应用形式成熟落地，既提高了用户对目标商品的获取效率，又改善了商家的投放效果。

数据管理与分析

大数据基础架构包括多种数据库及相关分析计算技术，能够支持数据的采集、清洗、集成、存取、分析、挖掘等环节，从功能角度来看，除传统的数据处理过程外，可视化分析等分析服务也是基础架构的一环。

- **应用情况：**企业（含互联网企业）的数据相关部门（如数据中心/数据分析平台/数据商业部等），主要通过数据的管理、分析及整合，服务于自有业务，同时不断拓展应用类型和范围。
- **价值体现：**大数据处理过程中流转的数据形态可划分为静态数据和动态数据，主要通过计算过程实现其价值。其中，模型算法为数据深度挖掘的核心，能够实现预测等功能。

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_21504

