

中国大数据发展调查报告 (2015 年)

中国信息通信研究院

2015 年 5 月

版权声明

本调查报告版权属于中国信息通信研究院，并受法律保护。转载、摘编或利用其它方式使用本调查报告文字或者观点的，应注明“来源：中国信息通信研究院”。违反上述声明者，本院将追究其相关法律责任。

前言

2014 年是中国企业级大数据市场的起步之年，经过近两年的探索与准备，国内一些电信和金融等行业用户开始实际部署大数据平台与应用，带动软件、硬件和服务市场发展。为进一步掌握中国企业大数据的应用状况和特点，中国信息通信研究院开展了 2014 年中国大数据应用状况的调查。本次活动采用在线问卷的方式，对中国大数据市场规模、应用状况、数据资源情况、平台建设情况、企业大数据应用的政策和资源需求等问题进行调研，调研对象包括电信、互联网、金融、批发零售等行业的固定从业人员 16368 人。本报告希望以严谨客观的数据，为政府、企业等了解中国大数据发展状况和制定相关决策提供参考。

本《报告》的数据采集工作得到了政府部门、大数据解决方案提供商和企业用户以及社会各界的大力支持。我们在此表示衷心的感谢！同时也对接受大数据调查访问的企业朋友表示最诚挚的谢意！

中国信息通信研究院

2015 年 5 月 29 日

目录

| | |
|-------------------------|-----|
| 前言..... | III |
| 目录..... | IV |
| 观点摘要 | 1 |
| 一、 调查背景..... | 3 |
| 1.1 调查方法及样本 | 3 |
| 1.2 报告术语界定..... | 4 |
| 二、 大数据市场规模..... | 5 |
| 2.1 大数据市场规模预估 | 5 |
| 2.2 大数据市场细分领域产值 | 6 |
| 三、 大数据应用 | 6 |
| 3.1 对大数据应用的认识 | 6 |
| 3.2 大数据应用的部署现状 | 7 |
| 3.3 大数据应用带来的效果 | 8 |
| 3.4 大数据应用的主要障碍..... | 8 |
| 四、 大数据资源 | 9 |
| 4.1 数据资源的规模..... | 9 |
| 4.2 数据资源的来源..... | 10 |
| 4.3 数据资源的类型 | 11 |
| 五、 大数据平台 | 12 |
| 5.1 大数据平台建设模式 | 12 |
| 5.2 自建大数据平台模式 | 12 |
| 5.3 采购公共云服务模式 | 15 |
| 六、 政策需求和资源需求 | 16 |
| 6.1 大数据发展的政策需求 | 16 |
| 6.2 企业对大数据资源的需求状况 | 17 |

观点摘要

(一) 我国大数据市场规模

- ✓ 2014 年我国大数据市场规模达到 84 亿元人民币，预计在 2015 年将达到 115.9 亿元，增速 38%。
- ✓ 大数据软件产值为 35.6 亿元，占市场总量的 42%，占比高于国际平均水平。

(二) 大数据应用现状

- ✓ 用户对大数据高度认同，调查中受访企业认为大数据技术对企业非常重要的超过 95%。
- ✓ 企业大数据应用效果明显。受访企业有 61.7% 的人认为提升运营效率是企业大数据应用的最明显效果，随后是实现智能决策（57.8%）和风险管理（45.6%）。
- ✓ 政策限制和数据资源短缺成为制约企业大数据发展的最主要因素。

(三) 企业大数据资源情况

- ✓ 大部分企业数据总量在 50TB~500TB 之间，占 73%。500TB 以上的只占 17%。
- ✓ 企业内部数据依然是开展大数据应用的基础。来自企业内部平台、企业客户和企业内部管理平台的数据占企业大数据应用资源总量的 52.9%。
- ✓ 企业传统结构化数据库表成为企业最主要的数据类型。73.9% 的企业认为数据库表是大数据应用当中最主要的数据类型。随后是网页数据（56.2%）和图片数据（34.3%）。

(四) 大数据平台建设模式

- ✓ 企业大数据应用的建设模式分为自建平台和购买云服务。超过 60% 的企业选择自建大数据平台，37% 的企业选择通过购买云服务的形式建设大数据平台。
- ✓ 企业大数据集群规模大部分比较小。受访企业用于大数据的服务器数量 50 台以下的占 49%，100 台以下的占 74%。

(五) 政策需求和资源需求

- ✓ 政府的各种大数据扶持政策成为企业关注焦点。71%的企业希望政府资助更多大数据领域的科研项目，其次是开放更多政府公共信息资源（64.7%）和促进数据流通交易（63.6%）。
- ✓ 企业对于政府公开数据需求强烈。42%的企业希望获得各类企业信息，随后是社保数据（27.9%）、教育科研数据（26.9%）和市政管理数据（25.5%）。

CAICT 中国信息通信研究院
China Academy of Information and Communications Technology

一、调查背景

1.1 调查方法及样本

本次调查采用问卷调查的方式针对大数据应用模式、大数据软硬件产品使用情况、大数据投入状况等问题进行了调研。

- 有效样本选取：

本次调研共获得 31000 份答题样本，选取其中的固定职业者 16368 份，根据行业分布和对大数据的了解程度对受访对象进行筛选，最终得到 809 份有效样本进行统计分析。

- 受访企业所在地：

覆盖我国华北、东北、华东、中南、西南、西北六大区域及海外地区。

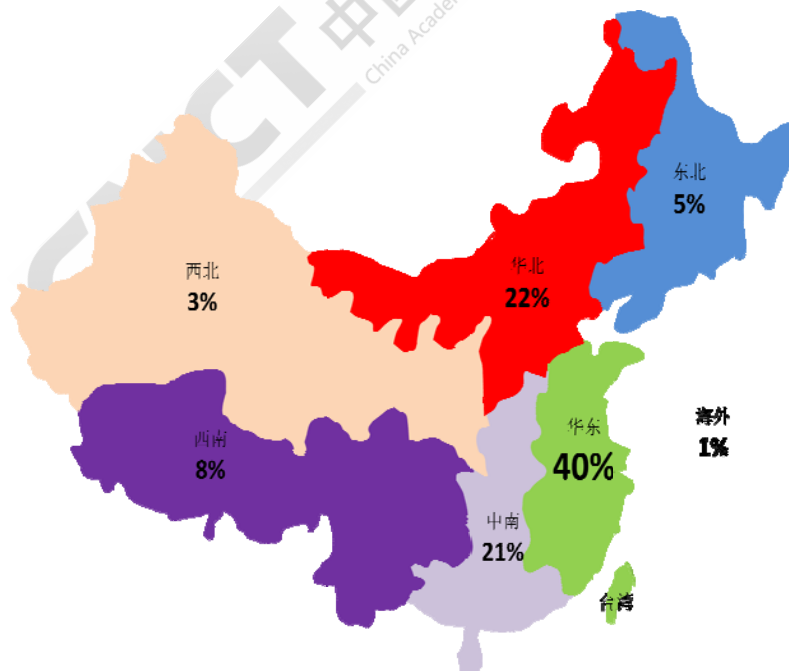


图 1 受访企业所在地区统计

● 受访企业规模和年盈利状况

大部分受访企业人员规模在 1000 人以内，200 人以下（21%）、200~500 人（29%），500~1000 人（20%）。

大部分受访企业年盈利在 1 亿元以内，500~2000 万元（25%）、2000~5000 万元（24%）、5000 万元~1 亿元（17%）。

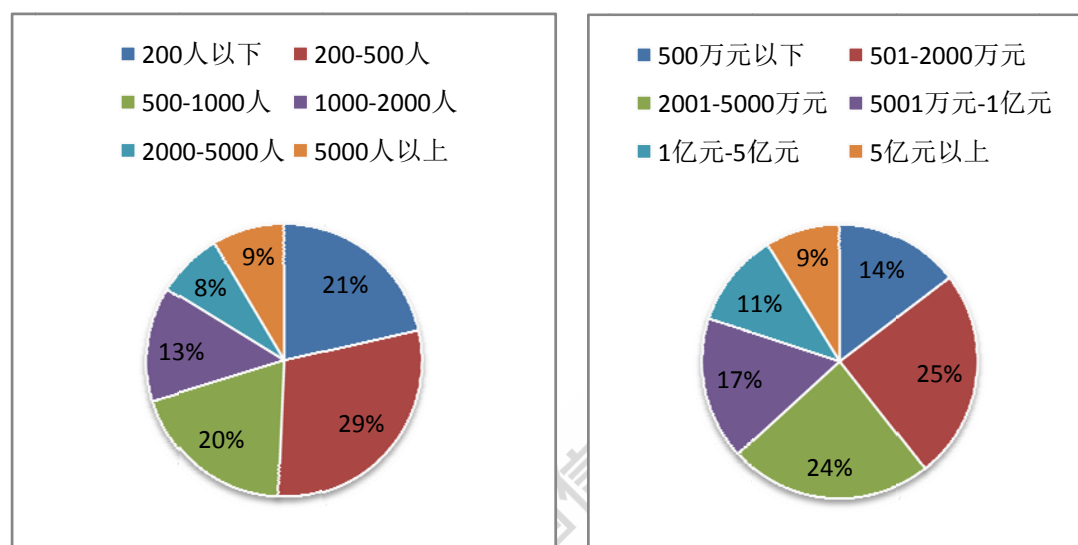


图 2 受访企业人员规模和年盈利情况

1.2 报告术语界定

大数据产业：本报告所称大数据产业，包含大数据直接相关的软件、硬件及专业服务。大数据市场规模的计算口径，相应的包含大数据软件、硬件及服务的直接产值，不含因大数据应用产生的附加值。

大数据软件：本报告所称大数据软件，主要指以 Hadoop 或 Spark 为基础的新型分布式数据平台软件，以及基于这类平台软件的应用软件。

大数据服务：本报告所称大数据服务，主要包括数据分析公共云服务，以及大数据系统相关的咨询、设计和系统集成服务等。

大数据硬件：本报告所称大数据硬件，是企业购买的指用于支撑在企业内部部署的大数据软件，主要包括服务器、网络设备和存储设备等。大数据系统一般使用通用硬件，技术上并无特殊性，也有少量专门优化过的一体机设备。

二、 大数据市场规模

2014 年，中国大数据市场规模约为 84 亿元，主要由基于 Hadoop、Spark 的大数据软件产值、用于承载大数据应用的硬件产值，以及大数据相关的专业服务产值三部分构成。

2.1 大数据市场规模预估

预计 2015 年中国大数据市场规模将达到 115.9 亿元，增速 38%。未来随着应用效果的逐步显现，一些成功案例将产生示范效应，预计 2016 至 2018 年中国大数据市场规模增长还将维持 40%左右的高增长。

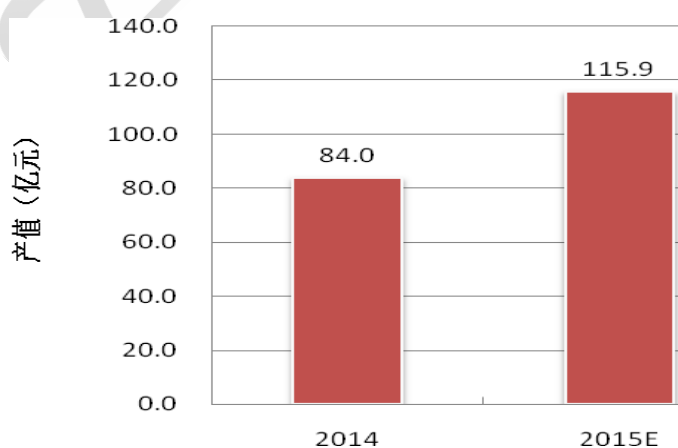


图 3 中国年大数据市场规模

2.2 大数据市场细分领域产值

从大数据市场构成来看，2014 年大数据软件产值为 35.6 亿元，占比最大，约 42%；大数据硬件产值 28.5 亿元，约占 34%；大数据服务产值为 19.9 亿元，约占 24%。与国外市场构成相比，国内软件占比明显偏高，这主要是由于国内客户更愿意接受软件版权费，而国外则更多的以收取服务费方式支付费用。

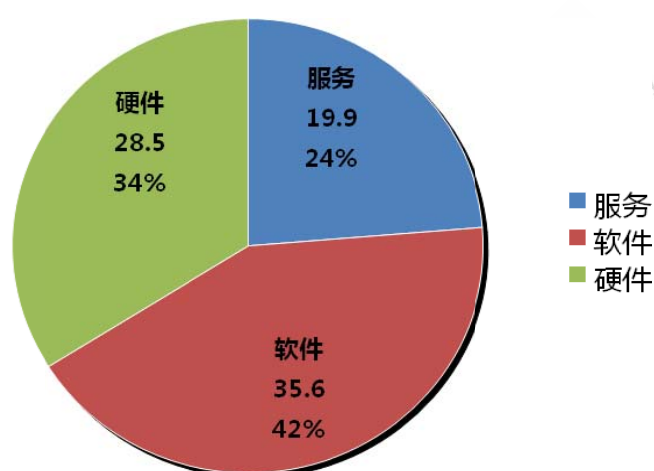


图 4 2014 年中国大数据市场细分领域产值

三、 大数据应用

3.1 对大数据应用的认识

调研结果显示，受访对象对于大数据应用能力对企业业务的重要程度认知基本一致，95%以上认为不可或缺或锦上添花。

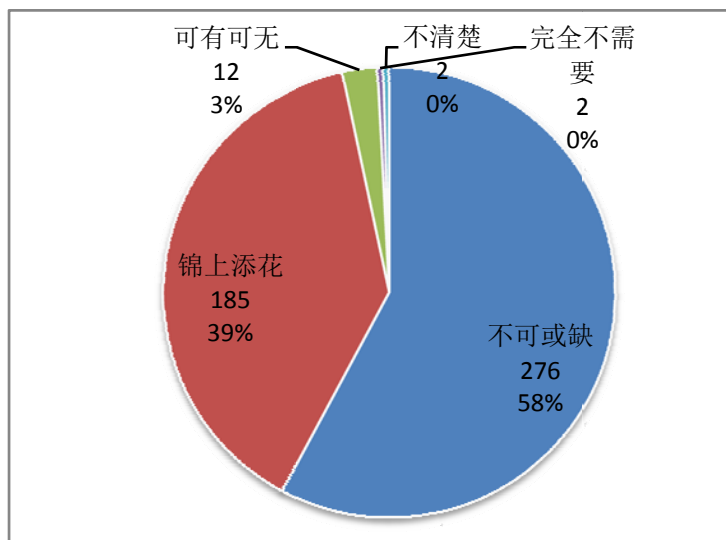


图 5 企业对于大数据应用认同程度

3.2 大数据应用的部署现状

调研结果显示，受访企业中有超过 44% 没有大数据平台部署和大数据应用，24% 的企业已经开始部署大数据平台但还未实现应用，32% 的企业已经实现了大数据应用。其中，大数据应用水平较高的行业主要分布在互联网、电信、金融行业，一些传统行业的大数据应用发展较为缓慢，批发零售业甚至有超过 80% 的企业并没有大数据应用计划。但较为乐观的是，在大数据应用变化情况调查中，95% 的企业大数据应用较上一年度呈增长态势。

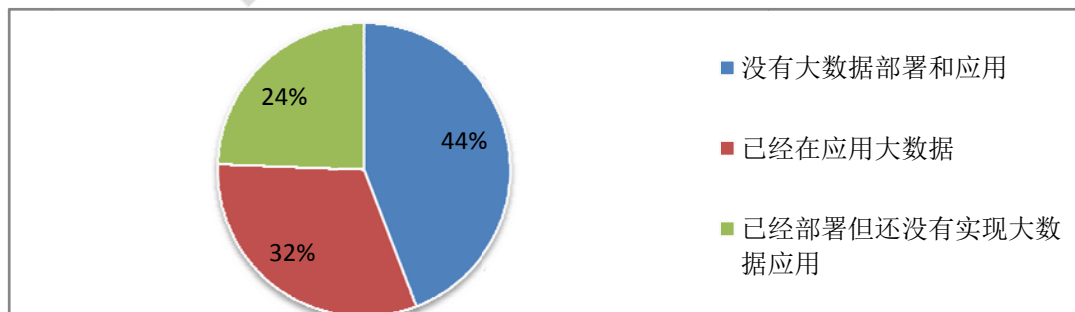


图 6 企业大数据应用状况

3.3 大数据应用带来的效果

在大数据应用效果方面，已经进行了大数据应用的受访企业一致将提升运营效率、实现智能决策和更好的管理风险视作最明显的应用效果。其次在降低运营成本、创造新的业务收入和提升客户满意度方面也有着不同程度的应用效果。统计结果表明企业大数据应用热点，也从另一个角度提示了企业在大数据应用的哪些方面需要加强。

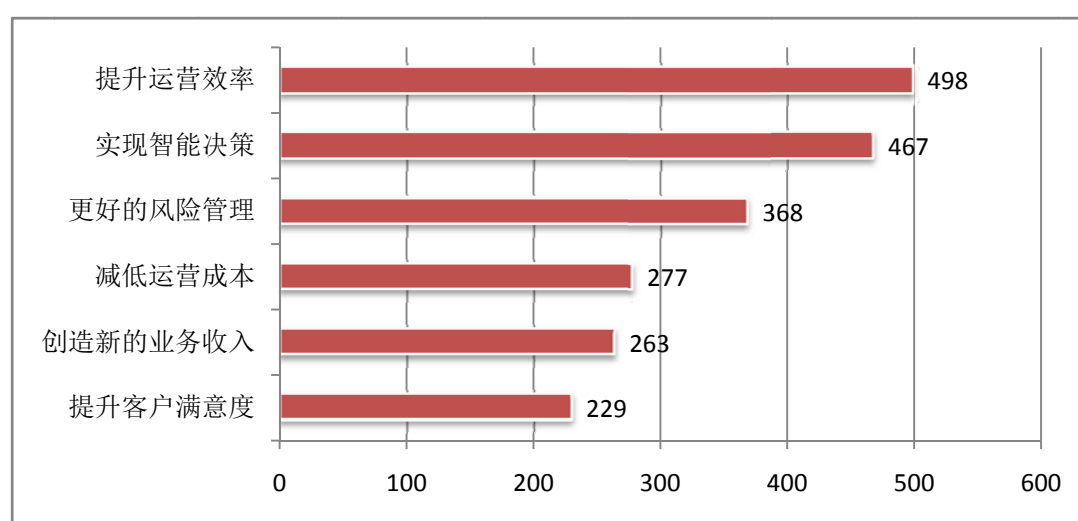


图 7 企业大数据应用效果（企业数量，n=809）

3.4 大数据应用的主要障碍

调研结果显示，制约企业大数据发展的最主要因素是政策限制（如隐私保护），其次是数据资源短缺和技术水平不足，随后是投资大和没有看到明显效益，最后是诸如人才缺乏等其他原因。大数据概念近两年热度不断上升，技术水平也飞速发展，但由于政策、法规的滞后性，数据跨部门、跨企业、跨行业甚至跨领域流动的需求被牢牢禁锢，这也使得政策限制成为了摆在我国企业大数据应用面前的最大障碍。数据资源增长迅速，但是如何通过技术手段获取高质量的数据

是企业面临的重要问题。大数据来源于互联网，对于一些信息化程度相对较低的行业来说，技术门槛是不得不正视的一个问题。很多企业基于投入产出比的考虑，认为投资大和无明显效益是大数据应用不能实行的的重要原因。

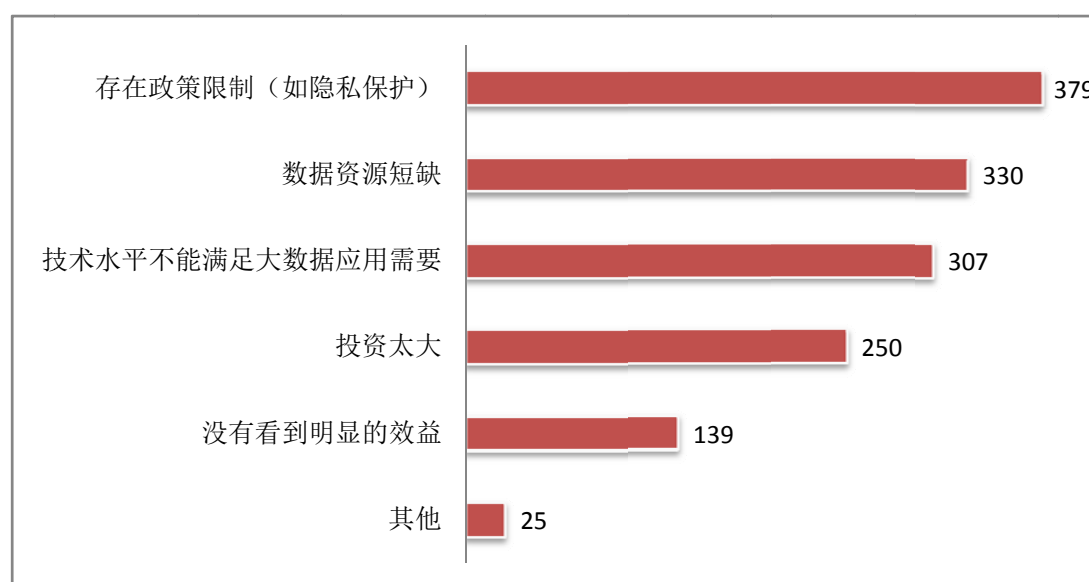


图 8 制约企业大数据发展主要因素（企业数量，n=809）

四、 大数据资源

4.1 数据资源的规模

在本次受访企业中，数据资源总量多集中在 50~500TB 量级上，总占比达到 73%，其中 50~100TB 数据量级占比最多（28%），低于 50TB 和高于 500TB 的企业只占 5%和 17%。随着企业数据的指数化增长，如何低成本高效率的采集、存储和利用这些数据资源，成为企业综合竞争能力的重要体现。

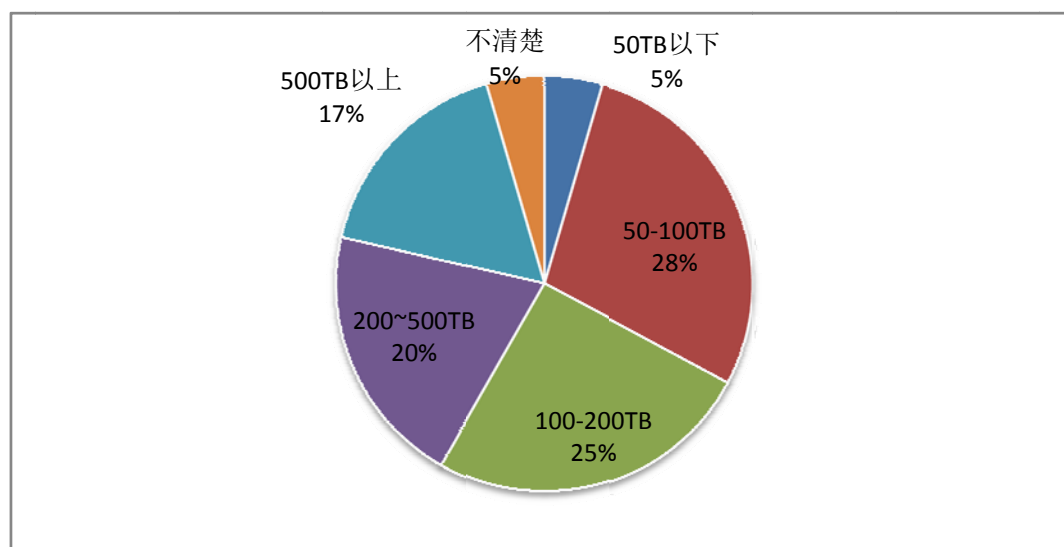


图 9 受访企业数据资源总量情况

4.2 数据资源的来源

调研结果表明，50%以上的企业把内部业务平台数据、客户数据和管理平台数据作为大数据应用最主要的数据来源。互联网上公开的数据和社交平台数据这类数据源，由于获取和分析技术不够成熟，价值密度低，分别有 40%和 35%的企业将这类数据纳入应用范围。32%的企业通过外部购买所获得的数据。只有 18%的企业使用政府开放数据。

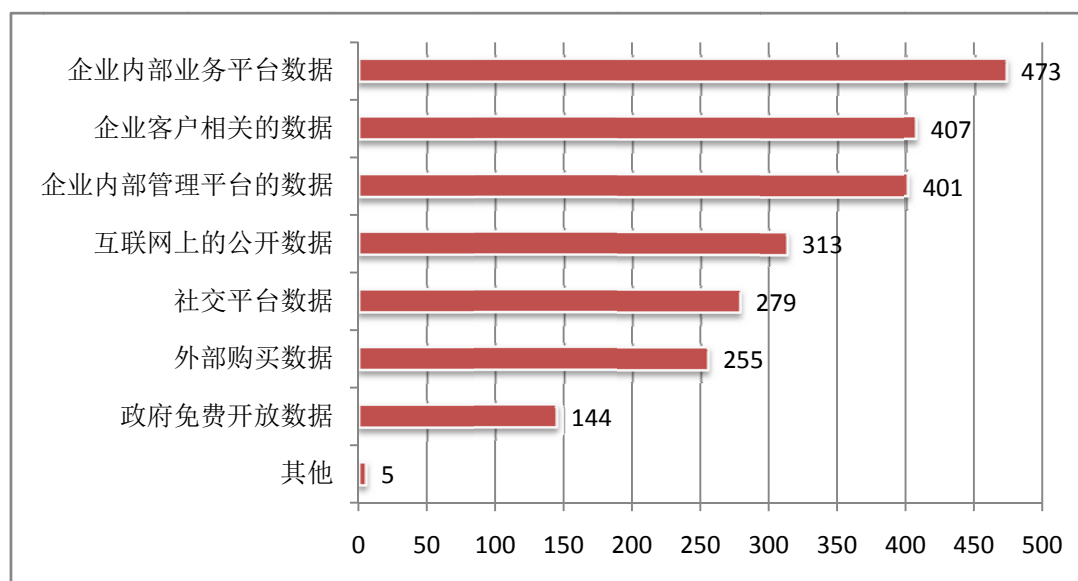


图 10 受访企业大数据来源情况（企业数量，n=809）

4.3 数据资源的类型

从数据类型上看，企业在实际应用中所使用的最主要的是已有数据库表，达到 75%，即关系型数据仍是企业的核心数据。其次是以网页数据为代表的半结构化数据，占 57%。由于分析技术发展相对缓慢，应用最少的数据来源是图片、语音、视频等非结构化数据。

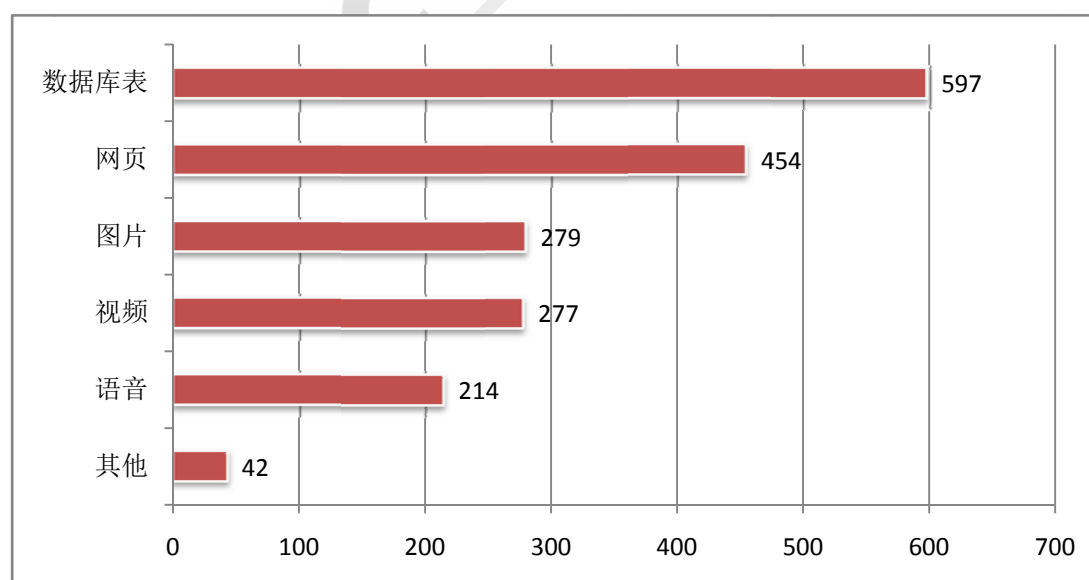


图 11 企业大数据应用类型（企业数量，n=809）

五、 大数据平台

5.1 大数据平台建设模式

本次调研根据大数据平台建设模式的不同分为自建大数据平台模式和采购公共云服务模式。自建大数据平台指的是购买大数据相关产品或解决方案，在本地建设大数据相关基础设施、存储和计算平台、分析工具和应用展示等。采购公共云服务指的是购买某一家大数据云服务提供商所提供的在线大数据服务。

调研结果表明，国内企业的大数据平台建设模式以自建私有云为主。有超过 60%的受访企业已经或计划通过以自建私有云的方式搭建大数据平台，只有 37%的企业已经或计划购买云服务的方式建设大数据平台。

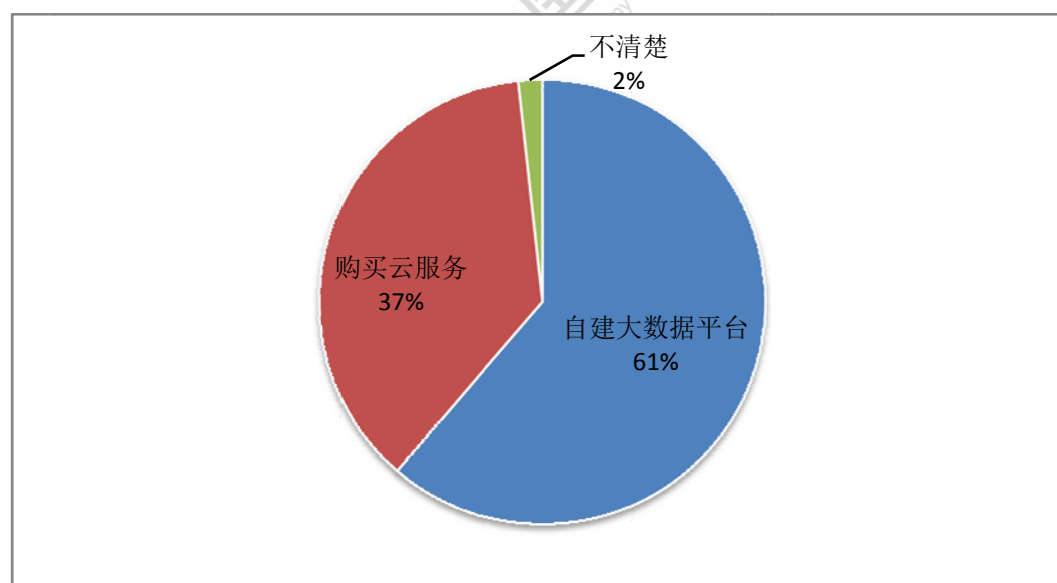


图 12 大数据平台建设模式分类

5.2 自建大数据平台模式

（1）大数据基础平台软件类型

闭源大数据商用软件仍是企业首选。在大数据平台软件类型选择方面，71%的受访企业选择使用闭源软件建设大数据应用平台。由于开源软件的使用和运维技术门槛较高，较少企业选择开源软件。只有18%的企业选择基于社区版进行自主二次开发，11%的企业选择基于开源商业版软件建设大数据应用平台。

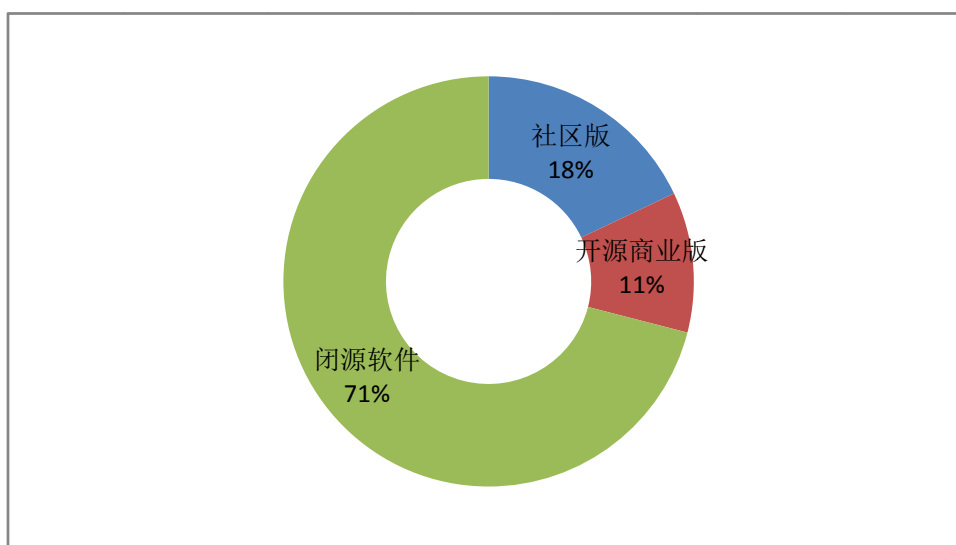


图 13 大数据平台软件类型

（2）服务器规模

受访企业中，用于大数据的服务器数量 50 台以下的占 49%，100 台以下的占 74%。100~500 台服务器和 500 台服务器以上的企业各占 13%左右。考虑到企业内部的服务器可能会分属于多个集群，因此企业中的大数据单集群规模在 100 台以上的应该远低于 20%。

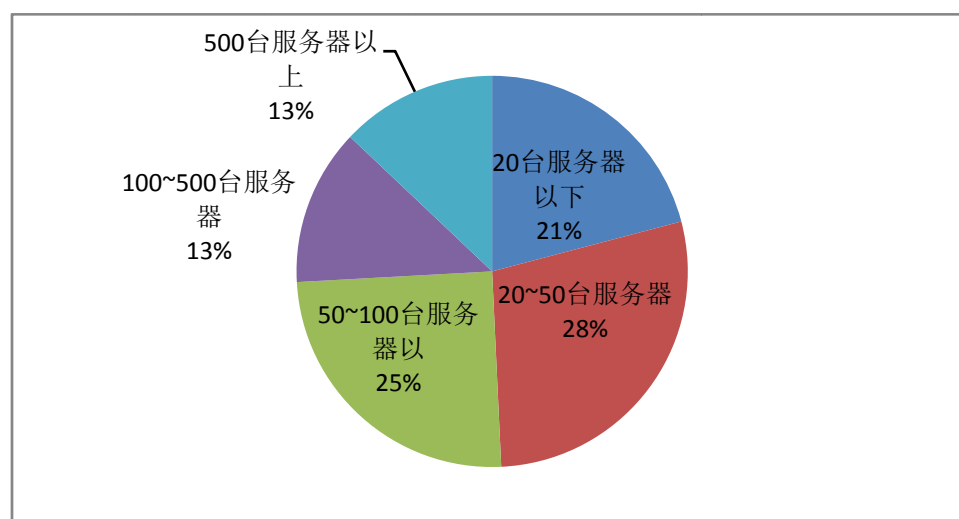


图 14 受访企业服务器数量情况

（3）选择大数据产品的考虑因素

受访企业在考虑选择大数据产品时都把产品的安全性放在首要地位（平均分 4.55 分，满分为 5 分），可见对于大数据产品的使用态度较为谨慎。接下来是产品稳定性（4.42 分）、售后服务的及时性（4.19 分）和产品易用性（4.14 分），相对来说放在次要考虑位置的是产品可扩展性（3.97 分）、产品性价比（3.95 分）、产品的成功案例（3.85 分）。

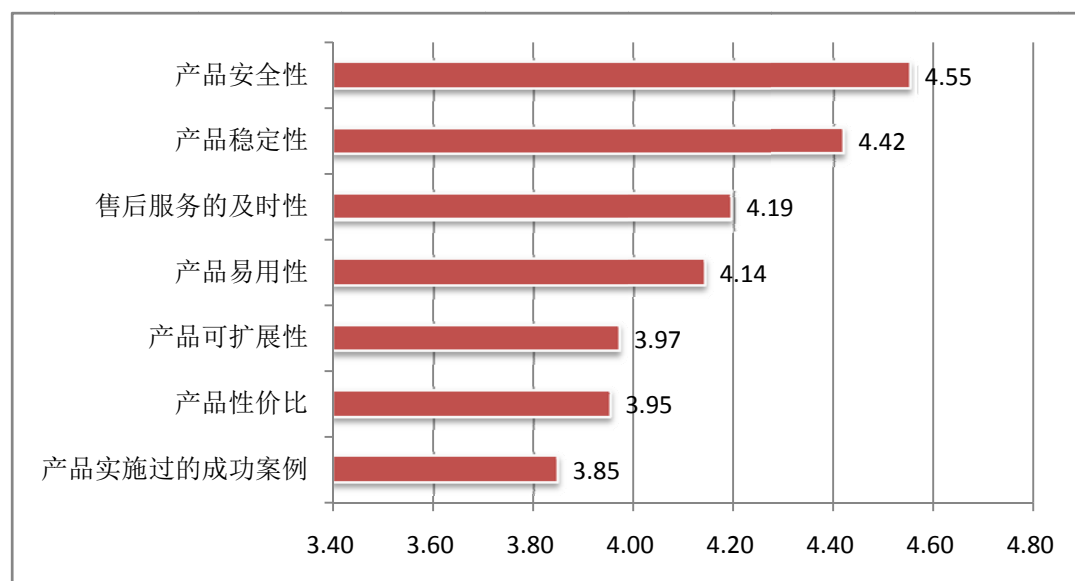


图 15 选择大数据产品时的考虑因素

5.3 采购公共云服务模式

(1) 云服务商选择

在被调研的近 800 家国内企业和 11 家海外企业中，由于国内云服务商性价比、易用程度和知名度不断上升等原因，有 88% 的企业选择使用国内云服务厂商提供的大数据云服务，只有 12% 的企业选择国外厂商提供的大数据云服务。

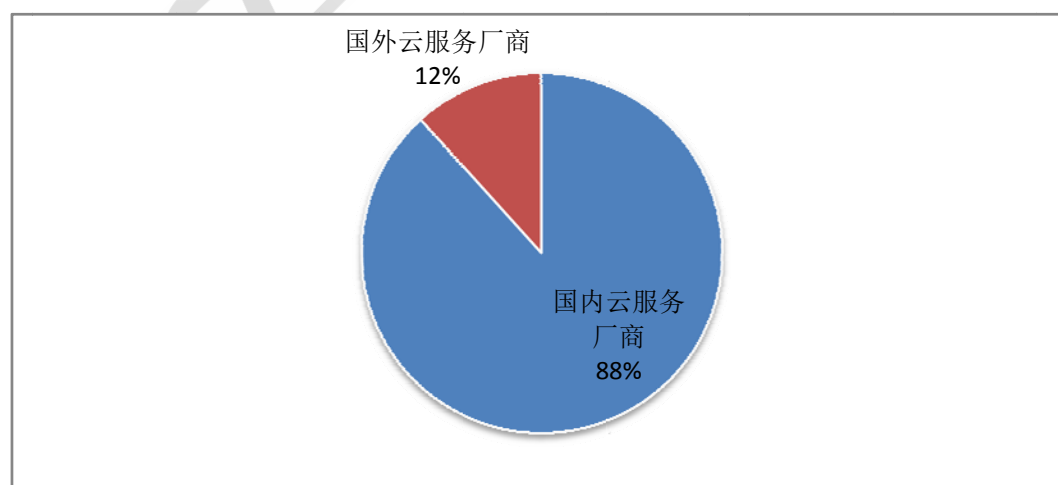


图 16 云服务购买选择

（2）云服务类型选择

在受访企业所购买的大数据云服务中，30%选择购买在线数据库产品，27%选择购买在线大数据分析工具，17%的人选择通用 IAAS（存储或虚拟机），12%的人选择在线 Mapreduce 服务。另外，企业对于可视化重视程度有所增加，13%的人选择在线可视化工具进行数据展示。

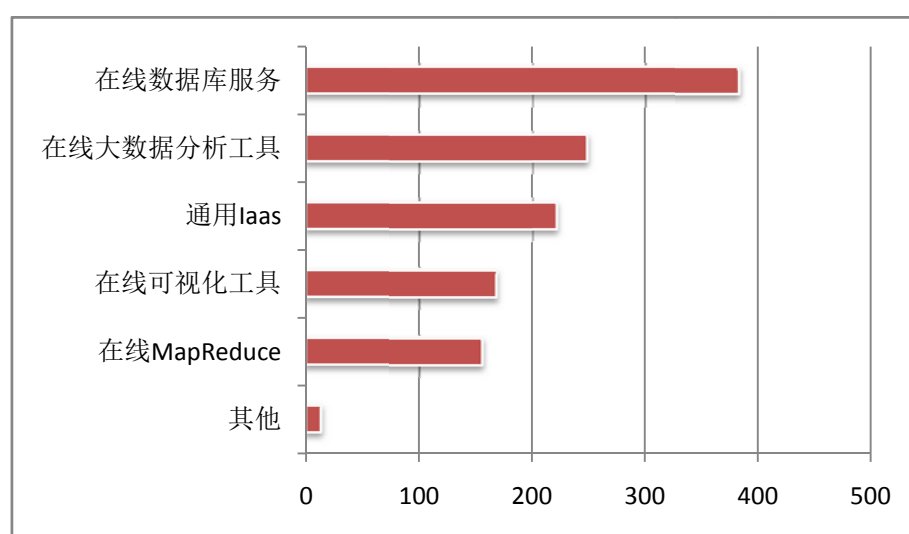


图 17 云服务购买类型

六、政策需求和资源需求

6.1 大数据发展的政策需求

政策环境对于大数据发展至关重要，企业对于各种政策的期望也有所侧重。最多的企业希望政府能够自主更多大数据领域的科研项目；随后是开放更多政府公开信息资源和促进数据交易流通；个人数据保护和加大政府采购力度也成为企业关心的因素之一。

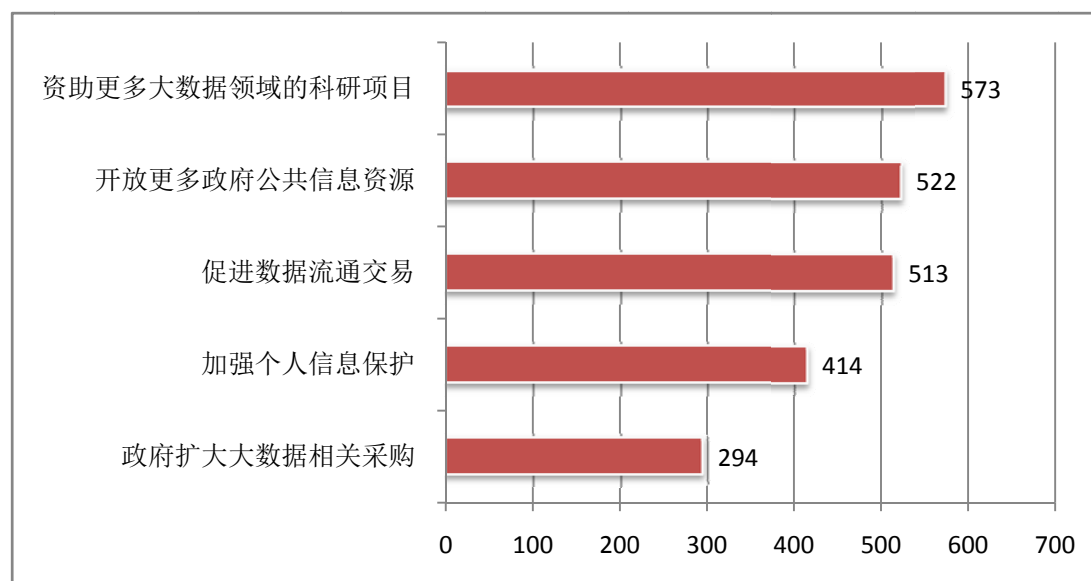


图 18 企业对于政府和政策的需求

6.2 企业对大数据资源的需求状况

企业对于大数据来源的领域有着不同期望，调研结果显示，企业信息公开成为最大诉求，随后是社会保障数据、教育科研数据、施政管理数据、劳动就业数据，一些专业数据的需求相对较少，包括交通服务、医疗健康、食品药品、气象环境等。

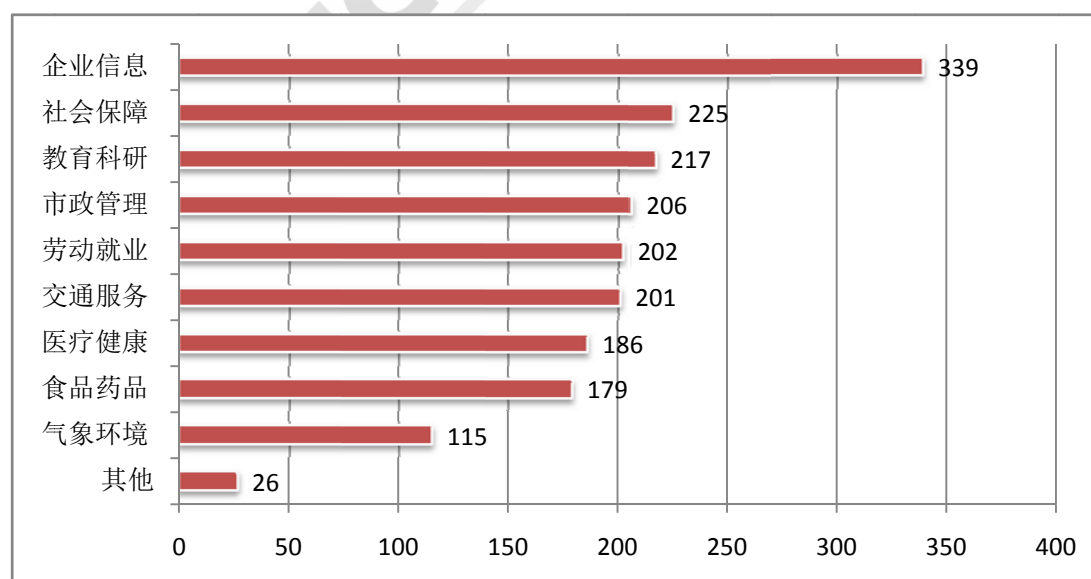


图 19 企业对大数据资源的需求