



# 数据时代 2025:

数据的不断发展，将对日常生活发挥出至关重要的作用  
不仅要关注大数据，还要关注重要的数据

## 概要

我们正在快速迈进数据时代的新纪元。从自动驾驶汽车到人形机器人，从智能个人助理到智能家居设备，我们周围的世界正在发生一种根本的变化，改变着我们的生活、工作和娱乐方式。

想象一下，虚拟的个人助理叫醒你，照顾你，并且根据天气预报和当天日程建议你最好穿什么衣服；然后安排自动驾驶汽车载你出门。或者，也许你根本不需要通勤去办公室，因为先进的技术使你能够利用交互式界面凭空“召唤”出工作空间，而全息视频会议则成为你与同事进行虚拟沟通的常规标准。周末，你可以通过增强现实应用来浏览新家具，看看某个沙发放在你的客厅是否合适，然后再下单。周六晚上，正当你坐在新买的沙发上放松时，无人机送来了外卖，是机器人制作的美味披萨。

过去 30 年，数据对人类生活的方方面面都发挥出了至关重要的作用：它改变了我们接受教育和享受娱乐的方式，它还告诉我们如何去体验繁荣的商业和更为广泛的世界。对于我们迅速发展的数字化存在而言，它可谓一条生命线。数字化存在即我们地球在任何一年时间内创建、捕捉和复制的全部数据总和。显然，它的发展极为迅速，而我们将其称为“全球数据圈”(Global Datasphere)。仅过去十年，我们的社会已经完成了从模拟到数字的转型。下一个十年，数据的力量必将给我们带来几乎无限的变化。

作为消费者，我们将享受到数字化存在的种种好处，而全球各地的企业也将迎来全新而独特的商机，这都要感谢海量的数据及其提供的洞察力。从海量数据的复杂性中提取并提供简单性和便利性——无论是通过机器人、3D 打印还是其他有待出现的技术创新——将会司空见惯。机会看上去是无限的，就像联网设备和服务所创建的庞大数据量一样。

从电网和供水系统到医院、公共交通及道路网络，实时数据的增长在数量和重要性上都显得极为引人注目。曾经主要推动企业成功运营的数据，现如今成为了消费者、政府和企业日常各方面顺利运作的关键要素。

在由希捷赞助的本白皮书中，IDC 分析了从现在到 2025 年推动全球数据圈发展的各种趋势。此外，我们还仔细考察了在管理、存储和保护最重要的数据方面，这些趋势对人们和企业所产生的影响。

IDC 预测，到 2025 年，全球数据圈将扩展至 163ZB (1ZB 等于 1 万亿 GB)，相当于 2016 年所产生 16.1ZB 数据的十倍。这些数据将带来独特的用户体验和众多全新的商业机会。

《数据时代 2025》指出了五大关键发展趋势，它们将强化数据在改变世界进程中的作用：

- **由商业数据向生活关键型数据的转化。** 曾经孤立、遥不可及、利用不足的数据，现在却成为了我们社会和个人生活的必需品。事实上，据 IDC 预计，到 2025 年，全球数据圈中近 20% 的数据将成为影响日常生活的关键数据，其中近 10% 将变为超关键数据。
- **嵌入式系统和物联网 (IoT)。** 联网的数字设备逐渐取代了独立的模拟设备，并将产生大量的数据，而这些数据反过来又让我们有机会以前无法想象的方式改良、改进自身的系统和流程。大数据和元数据 (数据的数据) 最终将触及我们生活中几乎每一个方面，并带来深远的影响。到 2025 年，全球平均每人每天与联网设备互动的次数将达到近 4800 次，基本上每隔 18 秒钟一次。

- **移动和实时数据。**无论人们何时何地需要数据，数据都需要即时可用，这一趋势越来越明显。目前，全球各行各业都在经历由这种需求推动的“数字化转型”。到 2025 年，在全球数据圈创建的数据中，超过四分之一的数据在本质上都是实时数据，而物联网实时数据将占这部分数据的 95% 以上。
- **认知/人工智能 (AI) 系统改变格局。**大量涌现的数据催生出了一系列全新的技术，例如：机器学习、自然语言处理和人工智能 (统称为认知系统)，它们将数据分析从不常见的、追溯式的实践转变成为战略决策和行动的前摄式推动因素。认知系统可以大大提高各种行业、环境和应用数据分析的频率、灵活性和即时性。IDC 预计，到 2025 年，属于数据分析的全球数据总量将增长至原来的 50 倍，达到 5.2ZB；而认知系统“触及”的分析数据总量将增长至原来的 100 倍，达到 1.4ZB！
- **安全成为一种重要的基础。**全新来源的各种数据将会给私人信息和敏感信息造成新的漏洞。如今，不断产生的大量数据需要安全保护，而实际得到安全保护的数据则少之又少，这之间存在显著的差距，而且这一差距还将继续扩大，这是数据驱动型世界的一个现实。到 2025 年，在全球数据圈创建的所有数据中，近 90% 的数据将需要某种程度的安全保护，但得到安全保护的数据将不到一半。

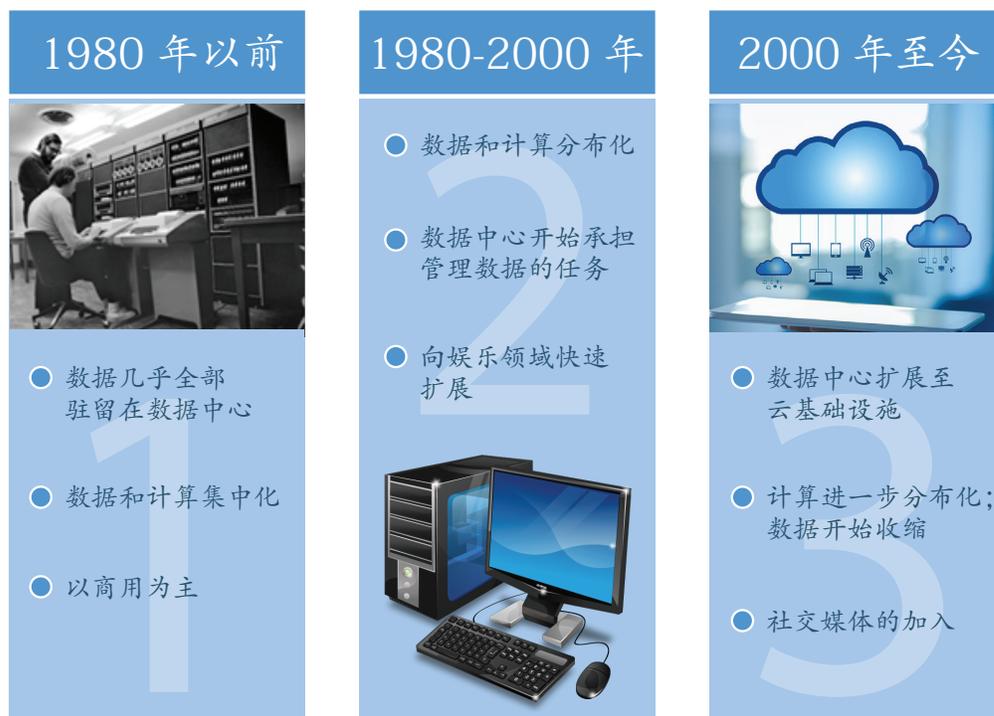
随着数据数量、种类和重要性的不断增加，企业领导者必须将注意力集中到最重要的数据上。并非所有的数据对于企业或消费者都同等重要。企业只有识别并充分利用至关重要的那部分数据，给用户体验带来有意义的积极影响，解决复杂的问题，并创造出全新的规模经济，才能够在数字化转型过程中实现蓬勃发展。企业领导者应当重视识别和维护这部分独特而关键的数据，以发挥其巨大的潜力。

## 由商业数据向生活关键型数据的转化

当代社会产生、使用和保有的数据量，即使不用“难以想象”一词来形容，按照以前的任何标准来说都可以称得上非常巨大。不仅如此，据 IDC 估计，未来几年全球数据圈的规模还将继续扩大，让现有的规模相形见绌。IDC 预计，到 2025 年，全世界将创建和复制 163ZB 的数据，是 2016 年所创建数据量的十倍。这种超速增长是过去数十年以来计算技术发展的必然结果。如图 1 所示，IDC 将计算数据的创建和使用大体分为三个主要时期：

- **第一阶段 (1980 年以前)**。在 1980 年以前，数据几乎全部驻留于专用的数据中心。即使人们通过远程终端访问数据，这些终端也是计算能力低下、甚至没有计算能力的非智能机。数据和处理能力仍然集中于大型主机。数据生成和使用的目的几乎完全是商用。
- **第二阶段 (1980-2000 年)**。个人电脑的兴起和摩尔定律的威力使得数据和计算能力更加大众化。数据中心从简单的数据容器发展成为集中化的枢纽，可以通过缓慢但不断发展的网络来管理数据，并将数据分发至终端设备。这些设备现在能够存储和管理纯粹由个人使用的数据，同时出现了音乐、电影和游戏的数字娱乐产业。
- **第三阶段 (2000 年至今)**。无线宽带和快速网络的普及推动数据进入云端，将数据与特定的物理设备脱钩，并开启了通过任意屏幕访问数据的时代。数据中心通过亚马逊、谷歌、微软等大众化服务扩展至云基础设施。随着手机、可穿戴设备、游戏机等新型设备的兴起，计算能力的分布化趋势不断发展。包括这些设备和传统 PC 在内的端点设备仍然需要数据进行操作，但是，通过云，可以轻易获取这些必要的的数据，本地存储的需求因此日益减少。这些趋势不断提升着 B2B、B2C 和社交互动领域中计算的重要性，反过来也受到它们的推动。

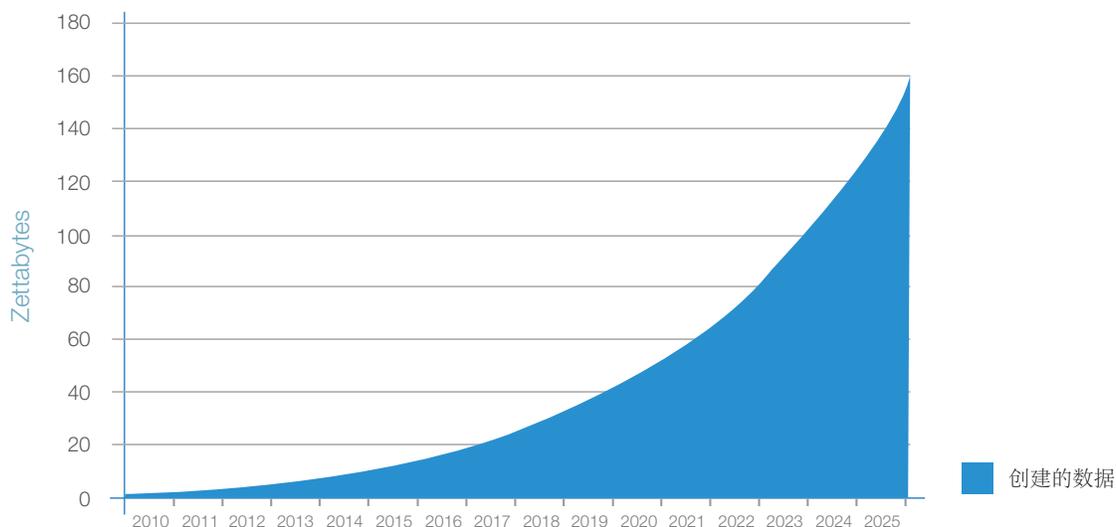
图 1. 计算的发展



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

这就是当今数据驱动型世界的现状。计算能力以及数据存储和可用性的巨大进步催生出了数字技术和服务的全新应用类型和场合。由此产生的需求反过来又推动着我们收集、管理、处理和交付数据的能力实现了进一步的发展，从而顺应企业工作流程和人们日常生活的需求。这个往复循环的结果使全球数据圈呈爆炸式增长 (见图 2)。

图 2. 全球数据圈的年度规模



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

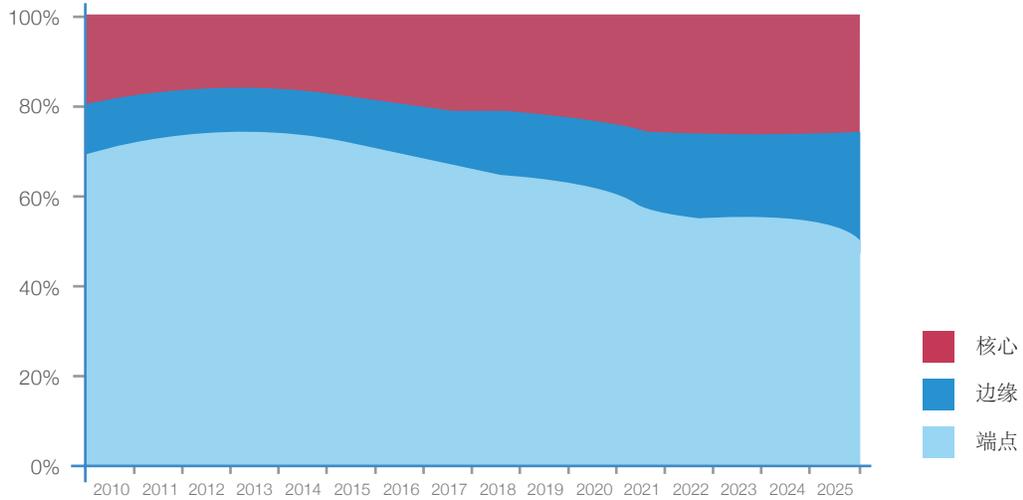
不同类型的计算平台在不同时间创建和使用的数据量，可以清晰地反应数据在这个世界上的演进作用。通过比较下列三种不同位置的计算平台，我们可以比较清楚地了解数据不断变化的作用：

- **核心**是指企业内部或云端的特定计算数据中心。它涵盖所有类型的云计算，包括公共云、私有云和混合云。此外，它还包括运营控制中心，例如管理电网或电话网络的那些运营控制中心。
- **边缘**是指不处于核心数据中心的企业级计算机/设备。它包括服务器机房，现场服务器，以及位于各个地区以加快响应速度为目的的小型数据中心。
- **端点**是指网络边缘的所有设备，包括PC、手机、摄像机、联网汽车、可穿戴设备以及各种传感器。

在总的的数据创建占比方面，端点自 2012 年以来大幅度降低，并且预计还将进一步下降 (见图 3)。过去十年，端点的增长主要来自于 PC、智能手机和其他消费类设备。尽管端点的增长仍将持续下去，但未来的增长将主要来自于嵌入式设备，包括安防摄像机、智能电表、芯片卡和自动售货机，这些设备产生的数据比较少。同时，大数据分析、云应用和实时数据的需求将推动核心和边缘平台以更快的速度增长。

虽然移动通信网络在速度和可靠性方面不断改进，但某些对时间敏感的应用关系到服务的质量甚至生命的维系，因此要求数据网络结构从数据中心核心外扩至动态的企业边缘。当实时数据和大数据分析发生交汇时，软件定义存储技术能够实现边缘存储环境的快速创建和迁移，从而满足本地及移动分析工作负载的需求。以这种方式提交数据，我们就能够实现联网汽车的无缝、高效交通流量管理(例如应急车辆优先交通协议)，或者在体育赛事或交通枢纽处开展实时欺诈检测或面部识别，从而确保安全。在移动、实时的世界，越来越多的联网设备生成越来越多的数据，这是推动边缘存储不断发展的基本要素。

图 3. 数据的创建位置

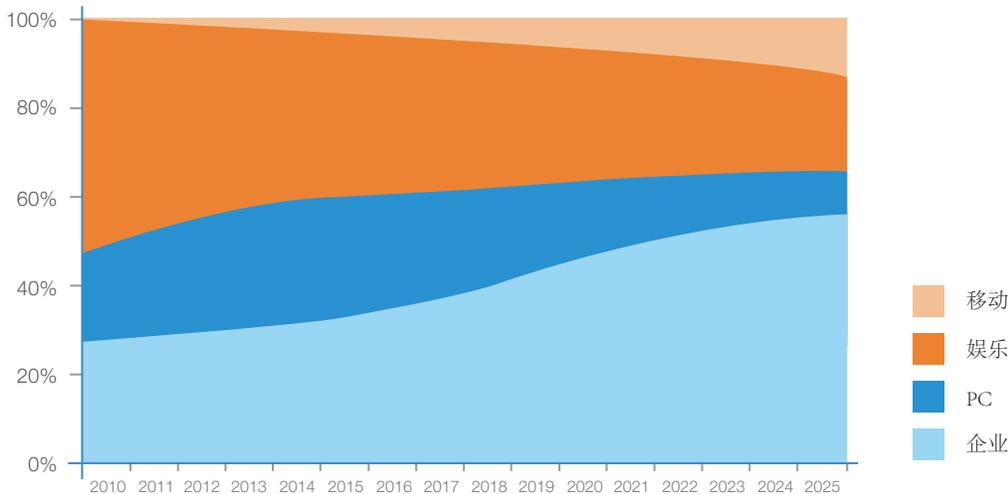


资料来源：IDC“数据时代2025”研究，希捷赞助，2017年3月

在评估生成并最终存储数据的平台时，我们还发现了另外一种快速变化的格局(见图4)。当前的基本情况之一就是企业再次成为数据使用的主要场合。从1980年到21世纪初，PC和娱乐媒体主导着数据的创建和消费。然而，随着时间的推移，网络和IP连接得到了改善，数据便不太需要以本地形式存储于PC或其他移动设备上。2010年，近50%的存储数据用于娱乐目的，因此DVD和蓝光光盘的销量非常大。之后，消费者的视频消费模式转向流媒体服务，企业基础架构存储的数据份额随之上升，而娱乐设备存储的数据开始下降。

其他变化则反映出了第三阶段的重要发展趋势，包括移动、社交媒体、大数据分析、高清视频和云计算。云存储的兴起增加了企业的数据使用。尽管移动设备规格较小，但其在考察的时间段内迅速崛起，原因在于企业竭力通过这些设备为客户提供实时的数据和服务。

图 4. 数据的存储位置



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

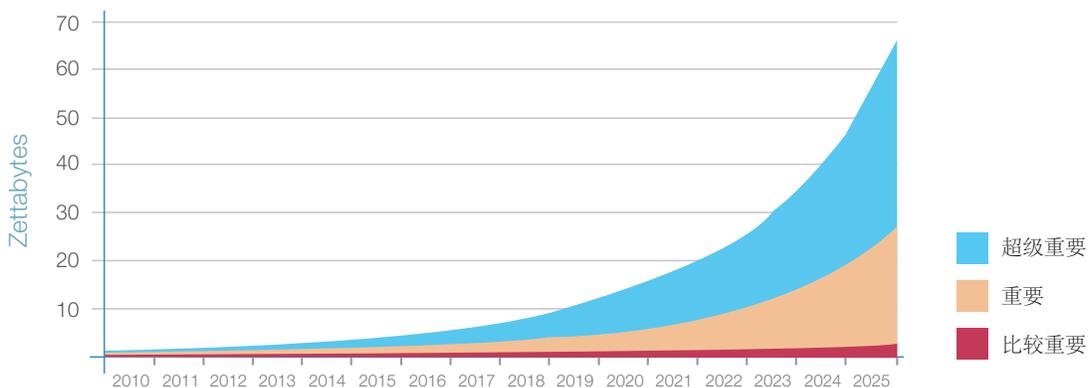
向基于云的、能够快速访问的、真正移动化的数据使用方式转变，其最终结果就是数据日益成为企业和日常生活各个方面的一项重要影响因素。

以商业航空旅行的现状为例：航空业周密部署各项资源，包括飞机、登机口、跑道、飞行机组人员和空中交通管制人员，以便发挥航空旅行基础设施的最大能力。这种相互依赖程度较高的系统其实比较脆弱，一旦系统的任何部分出现某个小问题，都可能引发多米诺骨牌效应，在数小时甚至数天之后、或者在数千英里之外扰乱正在进行中的旅行计划。航空业的应对之策是充分挖掘行程、延误、乘客人数、维护记录和天气方面的数据，以便预测潜在的问题并立即做出有效的应对。这些数据的使用方式一般较为传统(例如，将特定飞机作为资源进行分配时，参考航线的准时到达记录)，不过航空公司越来越多地以实时的方式来使用这些数据，以便随时根据具体情况实施调整。

数据的使用情况越来越多地按照其重要程度进行分析，而数据的重要程度体现在诸多方面，包括：实时处理和低时延的需求，使用的特殊性，以及数据不可用所造成后果的严重性（例如：一般认为医疗应用程序比流媒体电视节目重要）。IDC 估计，到 2025 年，全球数据圈中近 20% 的数据将成为影响日常生活的关键数据，其中近 10% 将变为超关键数据（见图 5）。

因为 PC 死机而丢失电子表格是一回事，因为自动驾驶汽车数据错误而造成人身伤害则是另外一回事。这样的事件不是关乎商誉的问题，而是关乎企业的生死存亡。超级重要数据的出现迫使企业不断开发和部署先进的数据捕捉分析技术和基础架构，从而提供极高的可靠性、带宽和可用性，实现更安全的系统，实施全新的业务实践，甚至采用全新的法律基础架构，以便降低转移和弱化责任的风险。

图 5. 数据的重要性



| 数据类型   | 2015-2025 年 CAGR |
|--|------------------|
| 所有数据。包括全球数据圈中的所有数据。  | 30%              |
| 比较关键。对于用户日常生活的顺利而言必不可少的数据。   | 37%              |
| 关键。对于用户日常生活的持续而言必不可少的数据。   | 39%              |
| 超关键。能够对用户的健康和幸福造成直接和即时影响的数据。(例如商业航空旅行、医疗应用、控制系统和遥测技术数据。这种类型的数据在元数据和嵌入式系统数据中的占比较大。) | 54%              |

资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

## 嵌入式系统和物联网 (IoT)

早期, 数据的增长主要源于个人电脑的兴起和数字娱乐的消费。现如今, 世界上的消费设备 (PC、手机、游戏机和音乐播放器) 比我们人类的数量都要多, 所有这些设备都需要数据进行操作。到目前为止, 我们已经大体完成从模拟电影和电视向数字化的转变。从 DVD 等离散型消费单元向流媒体服务的转变将继续推动数据的增长, 因为这个行业还在向着更高质量内容 (例如, 4K 或 8K 视频) 的方向发展。

计算能力嵌入大量的端点设备, 这成为了当前数据增长的主要因素。目前, 数据中心采用的嵌入式系统设备数量全球人均不到一台, 不过 10 年之后, 这一数字将增加到人均四台以上。虽然与娱乐及其他消费方式涉及的数据相比, 嵌入式系统产生的数据往往非常有效, 但其文件数量仍将非常庞大, 每年需以 1018 计。从这个数字的角度来讲, 尼亚加拉大瀑布需要 21 万年才能够喷流出 1018 加仑的水。

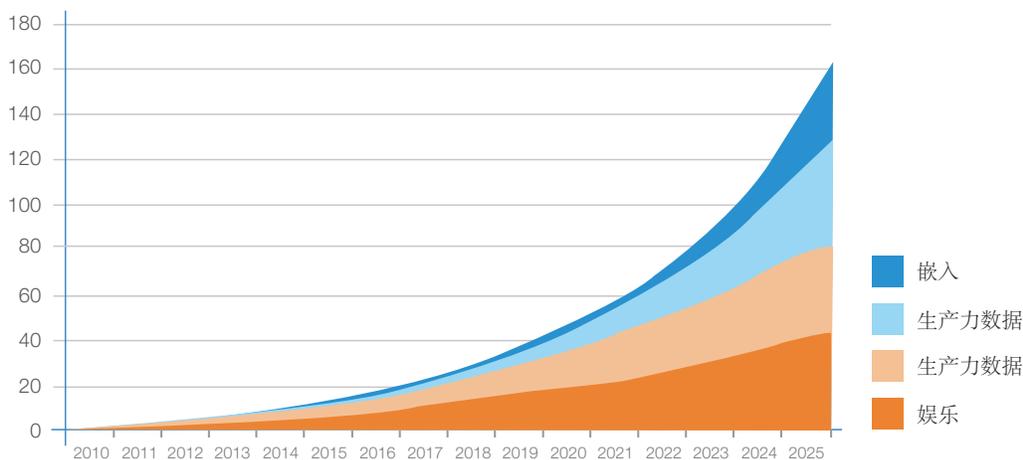
创建数据的所有这些嵌入式设备将推动大数据应用和元数据的增长, 并提升其价值。以元数据的应用为例, Netflix 非常善于利用视频观看数据: 通过监测用户的观看喜好 (例如: 喜爱的演员或影片类型), Netflix 能够给他们量身定制电影推荐列表, 以满足订阅者的实际需求。Netflix 的原创剧《纸牌屋》便是一个很好的例子。演员凯文·史派西 (Kevin Spacey)、导演大卫·芬奇 (David Fincher)、政治惊悚片和英国同名电视剧在 Netflix 用户中享有极高的人气, 最终促成了 Netflix 版本的诞生, 而其后来的成功也证实了这一做法的优点。

大多数嵌入式设备的数据不像 Netflix 列表那么直观, 但是这些设备还会生成运行方面的数据, 对它们所属的大型系统帮助巨大。像购物中心、交通网络和移动网络这样的系统会产生大量的原始数据点, 而这些原始数据点反过来又会生成关于它们自身的元数据。这种元数据不仅能够保证系统的持续运行和改进, 而且还有助于其他分析的背景判断。迪士尼主题公园的 MagicBand 对于游客来说非常实用, 因为它将公园通行证、房间钥匙和收费账户的功能非常简便地合而为一。此外, 它也是一种重要的数据来源, 能够帮助迪士尼优化其公园的运营并赚取利润。MagicBand 不仅能提供个人层面上的数据 (例如, 确定某人可以进入公园或打开房门), 而且它还能够实现各种各样的元数据分析, 从而了解公园游客如何行进, 如何利用公园及其附近的设施, 以及如何对迪士尼的营销刺激做出反应。

生成数据的设备类型多种多样，IDC 将全球数据圈划分为四大部分 (见图 6)。这四大数据创建类型分别是：

- **娱乐**。为娱乐目的而创建或消费的图像和视频内容。
- **非娱乐图像/视频**。为非娱乐目的而创建或消费的图像和视频内容，例如：视频监控画面和广告。
- **生产率数据**。传统的生产率驱动型数据，包括：PC 和服务器上的文件、日志文件和元数据。
- **嵌入**。嵌入式设备、M2M 以及 IoT 创建的数据。

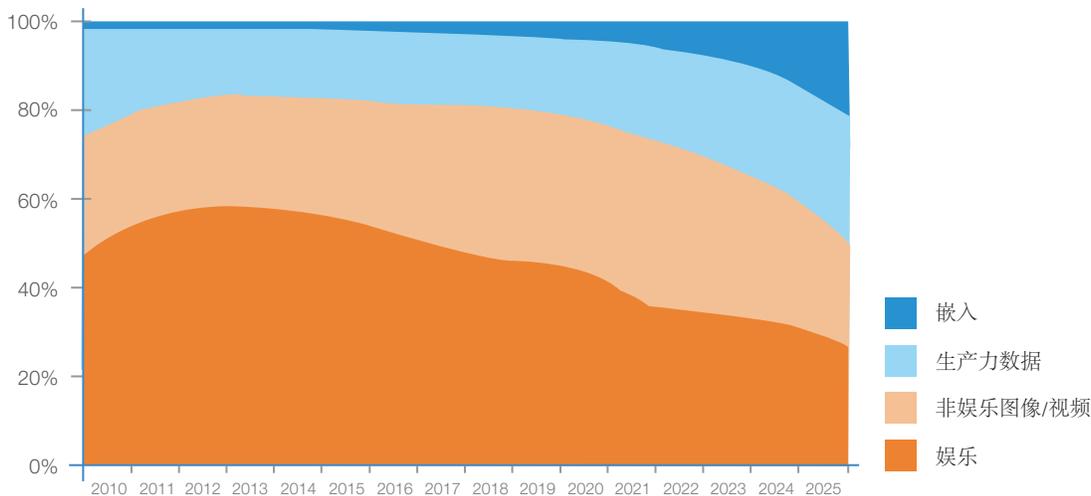
图 6. 数据创建的类型



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

数据创建的类型随着时间的推移而不断发生着变化(见图7)。通过分析数据创建类型的占比，我们发现，娱乐数据的比例急剧减少，而生产力数据、尤其是嵌入式数据不断崛起，将在我们未来的生活中发挥重要作用。

图 7. 数据创建类型的占比



资料来源：IDC“数据时代 2025”研究，希捷赞助，2017年3月

到 2025 年，嵌入式数据将占有所有数据的近 20%——为生产力数据的四分之三，并快速接近百分之百。生产力数据来源于传统的计算平台，如：PC、服务器、手机和平板电脑。另一方面，嵌入式数据则来源于极其广泛的设备类型，包括：

- 安防摄像机
- 智能电表
- 芯片卡
- RFID 读卡器
- 加油站
- 楼宇自动化
- 智能基础设施
- 机械工具
- 汽车、轮船、飞机、巴士和火车
- 自动售货机
- 数字标牌
- 娱乐场
- 可穿戴设备
- 医疗植入设备
- 玩具

所有这些嵌入式设备将大幅度提高人们与数据的平均互动水平，从而改变用户体验。这一趋势在类似 Facebook 这样的平台上已经非常明显，该平台可根据个人与特定内容类型互动的情况来调整内容和广告。未来 10 年，随着我们的家庭、工作场所、电器、车辆、可穿戴设备和植入设备具备数据功能，平均每人每天的数据驱动型互动次数预计将会提高 20 倍 (见图 8)。

图 8. 每人每天的互动



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

随着 AmazonEcho 等智能助手和汽车内置智能成为消费者日常互动环境 (平均每隔 18 秒互动一次) 的重要组成部分，绝大部分的互动将隐身成为背景。这种爆炸式的发展对于数据互动的最终影响将极其深刻，并导致社会和普通人日常生活的结构及质量发生不可逆转的变化。

尽管对日常生活有着深远的影响，但全球数据圈的绝大部分在使用之后都会被丢弃而非存储。这反映出了一个事实：大多数数据在使用或转移之后基本上都可以处理掉。回到之前的流媒体视频示例，我们没有必要为同一程序存储每个流媒体会话的内容。此时元数据可以发挥出作用。流媒体服务仅需保留特定视频观看事件的有关情况。这个情况可能会比较复杂，包括节目暂停或快进的时间和时长、观众是否看完整个节目以及通过什么设备观看。不过，元数据 (有益于流媒体服务的数据集) 要比原始流媒体事件小许多个数量级。这种方法充分吸取了过去十年数据增长的效率经验。我们必须从创建的海量数据中确定哪些数据具有足够的价值，哪些值得我们去存储。

同样，物联网设备也会产生大量经分析认定无需长期保留的数据。以视频监控摄像机为例。摄像机以视频的形式创建出了非常丰富的数据。通常情况下，大部分视频拍摄的是正常行为，没有保留的必要；只有小部分内容包含比较重要的事件，将来或许需要随时调用。在交通摄像机生成的数据中，地方交通管理部门重视的是交通违法或交通异常方面的视频，而会在创建适当的元数据之后丢弃正常、合法交通方面的视频。对于娱乐场所的视频监控系统来说，运营商仅仅重视和保留含有可疑行为的视频，其余部分在创建好元数据和经过一段时间后也会被丢弃。

在上述两个例子中，我们发现了智能标准的应用，也就是说，哪些数据需要保留、保留方式以及保留时间。这样，我们就可以着力于重要的信息，而不必存储所有生成的数据。这种有选择的数据保留策略标志着当前数据保留的最佳实践。

因此，数据生成的数量可以并且仍将继续超越我们存储所有数据的预期能力。例如，2025 年预计将创建 163ZB 的数据，而要想存储这些数据，我们需要大约 160 亿块套目前最大容量的 12TB 企业级 HDD。从这个角度来讲，过去 20 年中，硬盘行业共出货 80 亿块 HDD，总容量接近 4ZB。

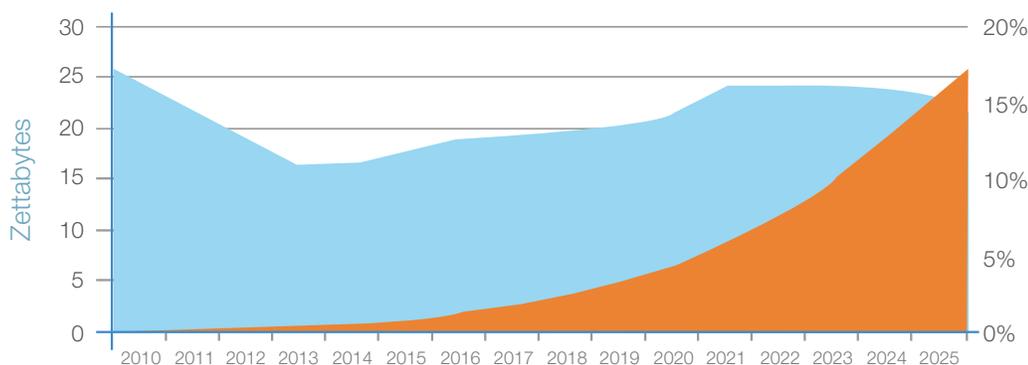
当然，我们有大量机会来存储更多的数据，无论这些机会是来自推动全球数据圈数据标记暴增的大数据应用，还是来自全新的数据保留管理规定。无论如何，根据目前的预期，数据存储需求都将继续保持迅猛增长的势头，而没有任何停止的迹象。IDC 预计，要想跟上“数据时代 2025”的预测，未来四年 (2017-2020 年) 出货的所有介质类型 (HDD、闪存、磁带、光学和 DRAM) 的存储器容量必须超过过去 10 年的出货量总和，即 5.5ZB。实际上，“数据时代 2025”研究预计，从 2017 年到 2025 年，所有介质类型的存储器出货容量都必须超过 19ZB，才能够满足存储需求。其中，大约 58% 的容量来自于 HDD，而 30% 来自于闪存技术。

## 移动和实时数据

联网的增加促进了移动接入和实时响应的发展。2005-2015 年期间，全球联网人数增长了五倍。同期，手机使用量超过了基于 PC 的互联网使用量，尤其是在几乎或者完全没有实体互联网基础设施的那些地区。到 2025 年，联网用户将占世界人口的 75%，其中包括以前没有联网的各种群体，如：儿童、老年人和新兴市场的人们。

未来几年,移动数据 (图 8) 和实时数据 (图 9) 都将呈现出强劲增长的势头。移动数据将维持自身在数据创建中的占比，而实时数据在总体数据创建中的占比将增长 1.5 倍。实时数据的使用可能 (但并非必然) 涉及到移动设备。例如，生产车间的自动化机器虽然固定不动，但也有赖于实时数据以实施工艺控制和改进。实际上，绝大多数实时数据的使用将由物联网设备驱动 (图 10)。

图 9. 移动数据



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

■ 占整个全球数据圈的 %

■ 移动数据

图 10. 实时数据

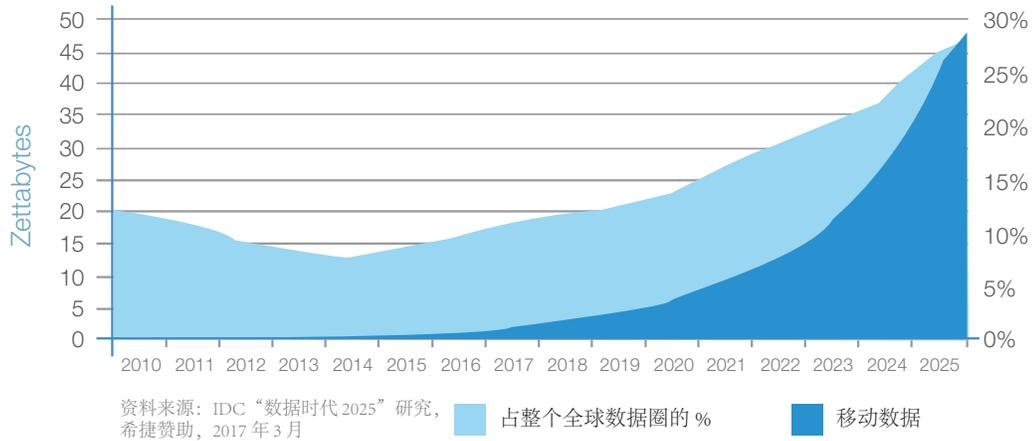
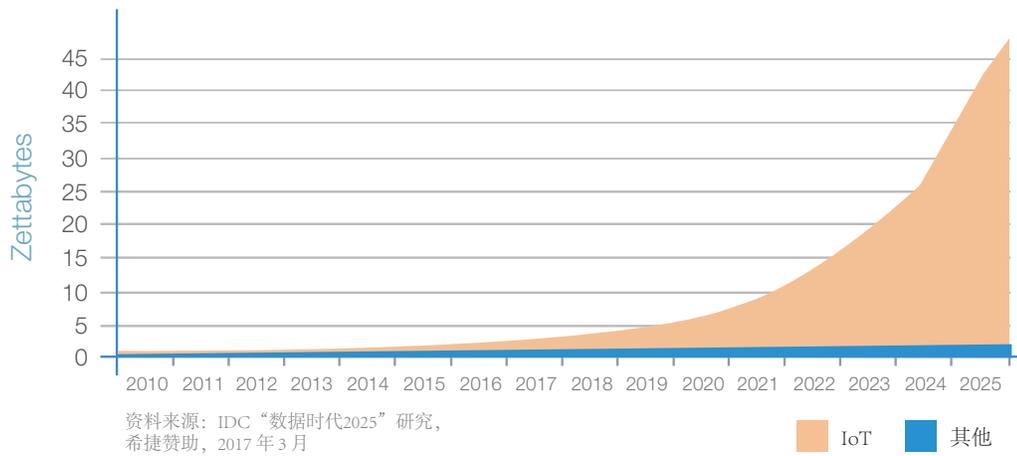
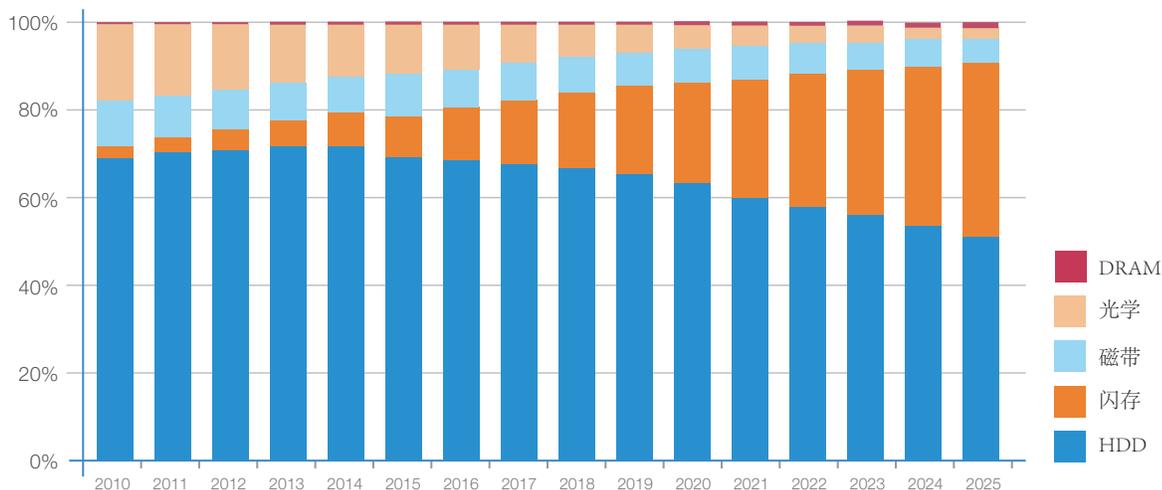


图 11. IoT 驱动的实时数据



实时数据的增长将导致未来数字存储类型的转变 (见图 12)。数据越来越多地要求以实时的方式提供，因此企业边缘存储和端点本身的低时延响应将日益受到重视。

图 12. 各种存储介质类型的出货占比



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

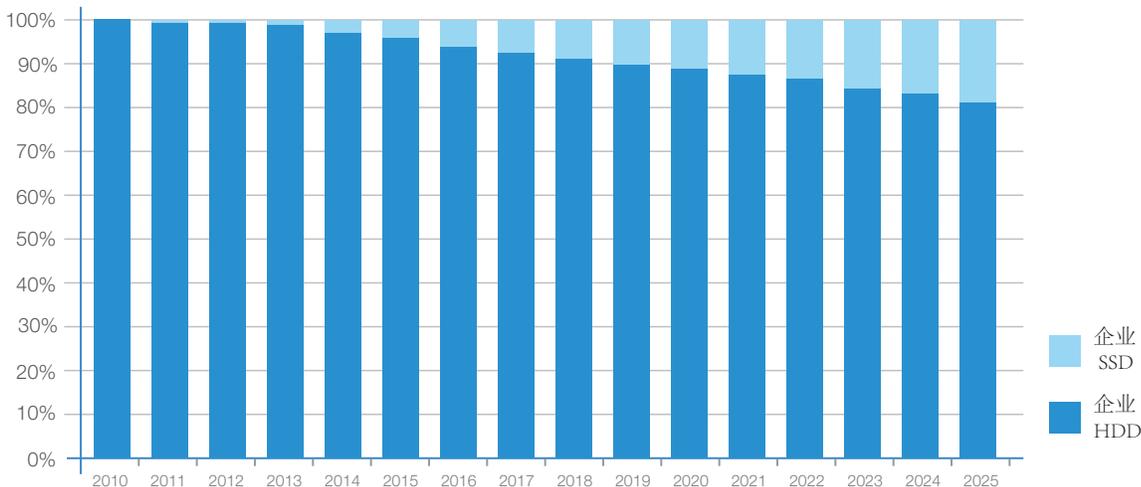
NAND 闪存的迅猛增长主要源于光学介质的没落。消费者对 CD 和 DVD 的利用远远少于前几年，他们现在主要通过流媒体网络来消费音乐和电影，光学介质因此变得不再那么重要。

随着实时数据的增长和联网用户数量的增加，企业数据与控制中心存储或“锚定”的数据量也在稳步增加，从而进一步扩展全球数据圈，其中有许多中心都将基于云。

实际上，据 IDC 估计，从 2016 年到 2025 年，公共云数据中心处理、存储或交付的数据百分比将近乎翻番，达到 26%。这些云可以处理、存储或提供的不仅仅有 IT 服务，而且还有娱乐、网络遥测和电信服务。

企业数据中心使用各种类型的存储介质，包括 HDD 和基于 NAND 闪存的存储设备(包括类似于闪存的新兴存储技术)，每一种存储介质都可以发挥出重要的作用，为广泛的存储工作负载提供经济有效的支持(见图 13)。

图 13. 企业出货：HDD 和 SSD



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

在一定程度上，磁带和光学存储还将继续作为传统的存储介质类型被企业数据中心所使用，但主要用于归档的数据，或者访问量非常少的数据。

### 人工智能系统改变格局

数据数量和可用性呈爆发式增长，这增强了认知/人工智能系统可为部署者提供的杠杆作用。

据 IDC 估计，到 2025 年，全球三分之二的金融公司将整合第三方认知数据，以便通过有针对性的产品和服务，提升客户体验和实现欺诈保护。这些认知系统的应用将触及企业和个人生活的方方面面。例如：

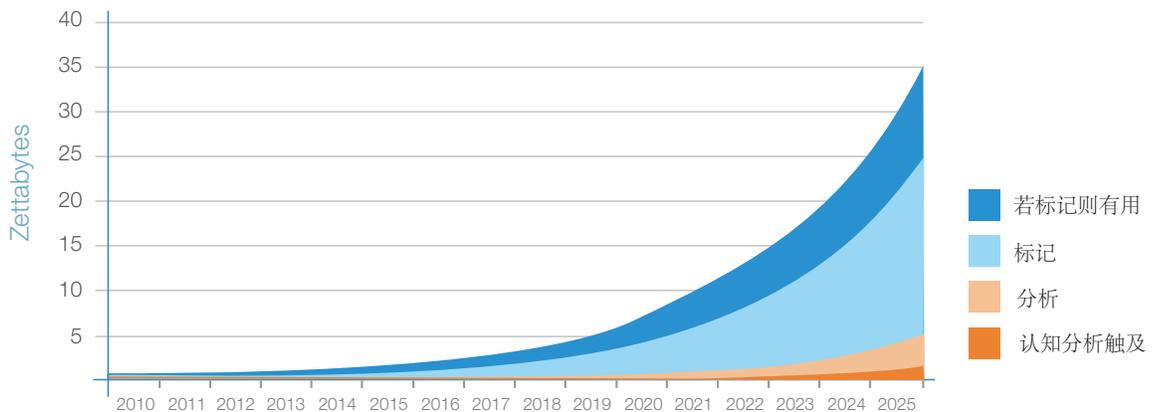
- 已然出现在某些城市街道上的无人驾驶汽车，有赖于实时遥测和机器学习来“学会”如何开车。这些基本认知系统的发展进步将缩短“教会”无人驾驶汽车如何开车所需的时间。
- AIG 和日本 Fukoku Mutual 等保险公司一直在使用基于人工智能的“代理人”和“虚拟工程师”，来处理索赔业务，提升工作效率。

- IBM 的 Watson 认知平台利用自然语言处理和机器学习等工具，帮助肿瘤专家和 美国 Memorial Sloan Kettering 中心开发针对性的个性化癌症治疗技术。
- 目前正在迪斯尼游轮上使用的面部识别功能可以实现所谓的“魔法艺术”，当 乘客走过时能够播放动画。该系统通过面部识别来确保乘客在随后的游览中不 会看到相同的画面。
- 包括 MasterCard 在内的大多数信用卡公司一般都通过人工智能进行欺诈检测。 这样他们就能够在短短 40-60 毫秒时间内检测出欺诈性交易。

数据标记 (特别是自动标记) 是认知系统应用的一个重要方面。归根结底,标记是给信 息添加标识符,以便于分类、分析、扩容和创造价值。不过,数据标记尚处于早期 发展阶段,还需要行业标准、更多的投资、更好的行业技术以及更多的数据科学家 参与 (见图 14)。尽管标记的数据并非总会受到重视,但标记数据的实际数量以及能 够受益于标记的数据数量之间仍然存在 (并且还将继续存在) 巨大的差距。

如图 14 所示,据 IDC 估计,到 2025 年底,全球数据圈中只有 15% 的数据将会被标 记,而只有五分之一的数据将会得到实际分析。

图 14. 数据标记



资料来源: IDC “数据时代 2025” 研究, 希捷赞助, 2017 年 3 月

不过，通过认知/人工智能技术还可以进一步挖掘自动数据标记的潜力。虽然这种方法尚处于形成阶段，但许多数据集成工具和系统正在构建其认知/人工智能功能，以便通过各种类型的机器学习（包括受监督学习、无监督学习和强化学习）将数据标记过程自动化。

## 安全成为一种重要的基础

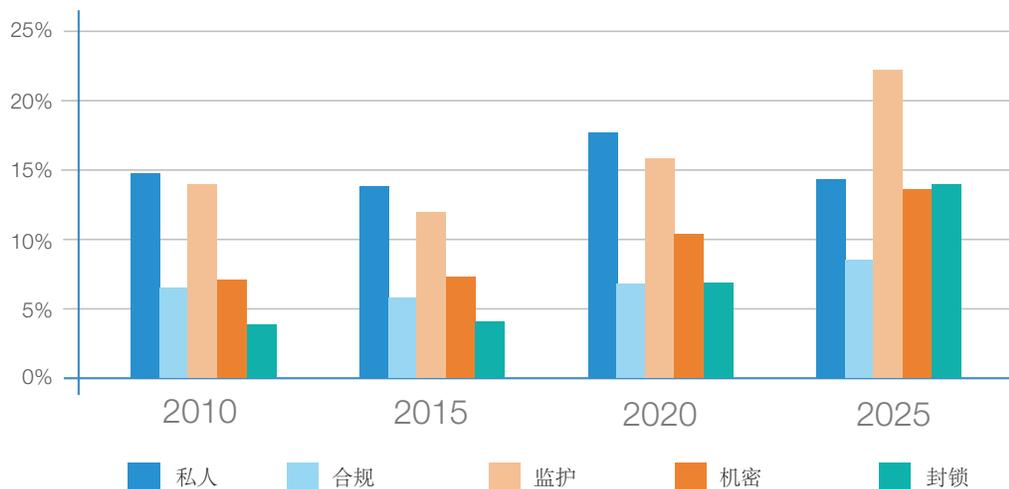
随着数据的来源、用法和价值不断发生变化，数据创建的数量也开始从消费驱动转为企业驱动。2015 年，企业创建的数据不到 30%，但这一数字将在 2025 年达到近 60%。（注意：在 1980 年以前，几乎所有的数据都由企业创建和管理）。无论数据在哪里创建，企业都必须面临管理全球数据圈中 97% 以上数据的严峻挑战。以社交媒体用户生成的内容为例：尽管上传个人视频和照片并撰写文字内容的是个人，但社交媒体网站最终必须在自己的基础架构中存储和管理这些数据。企业获得和管理的此类个人数据越来越多，其在管理隐私和安全风险方面承担的责任也越来越大。

此外，随着嵌入式传感器数量的不断增加，人们开始从不经意的交易中挖掘数据，额外的个人数据捕捉开始兴起。因此，数据安全的要求进一步提高。

部分类型的数据目前还没有硬性的安全要求，包括手机照片、数字视频流、公共网站内容和开源数据。然而，大多数数据都有安全要求，例如：公司财务数据、个人身份信息 (PII) 和医疗记录。到 2025 年，需要安全保障的数据百分比将接近 90%，这些数据可以分为五大类 (见图 15)：

- **封锁。** 需要最高安全保障的信息，如：财务交易、人员档案、医疗记录和军事情报
- **机密。** 原创者希望保护的信息，如：商业秘密、客户名单和秘密备忘录
- **监护。** 账户信息，一旦被突破，就有可能导致身份遭到盗用
- **合规。** 可用于诉讼或符合保留规则的电子邮件等信息
- **私人。** 如 YouTube 上传的电子邮件地址等信息

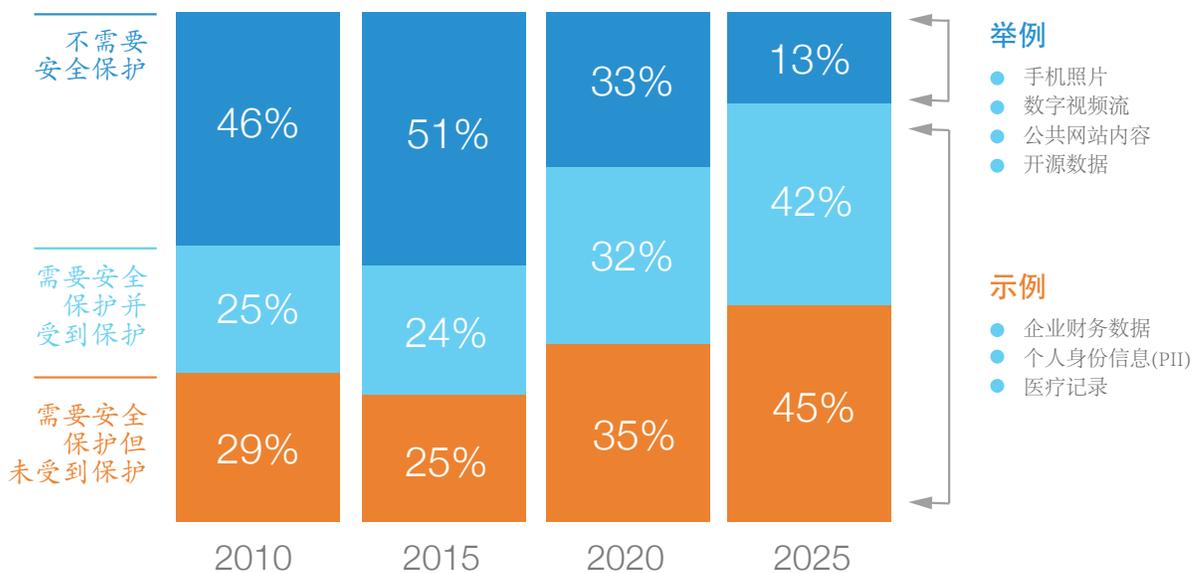
图 15. 需要安全保护的数据



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

令人惊讶的是，尽管绝大多数数据至少需要某种形式的安全保护，但实际上数据保护工作做得还远远不够 (参见图 16)。这种差距向安全与隐私技术、系统及流程解决这个问题提出了明确的行业需求。

图 16. 数据安全的实际状况



资料来源：IDC “数据时代 2025” 研究，希捷赞助，2017 年 3 月

## 结论

数据将对整个人类社会产生巨大的积极影响。数据不仅使得企业运营更加有效，而且正在改变个人生活的方方面面。不仅像 Uber 和 Netflix 这样的新型服务商有赖于数据，我们的城市、医院、商店、各类企业以及人类社会的方方面面也都有赖于数据。我们不断想方设法，让数据能够更好地服务于我们的生活，这在几年前简直难以想象。

社会使用数据的方式正在经历根本性的转变：

- 从娱乐转变为生产
- 从以商业为核心转变为超级个性化
- 从结构化转变为非结构化
- 从选择性转变为普遍性
- 从可追溯性转变为此时此刻
- 从提升生活品质转变为对生活至关重要

随着计算能力日益分布化，不断进入云端以及我们周围的日常物联网设备和基础设施，数据将为企业、行业、流程和日常生活带来根本性的改善。这些趋势将导致地球上的数据总量（即全球数据圈）呈指数式增长。全世界四分之三的人口将很快联网，数字化数据将影响到几乎每个人的生活，最终成为我们日益增长的数字化存在的生命线。

数据在企业及我们生活中的使用和整合正在迅速地实时化。因此，数据的交付不仅可以起到通知的作用，而且还可以帮助我们确定采取何种行动——有时候，这一过程甚至可以自动完成。虽然娱乐仍然是数据创建和消费的重要推动因素，但生产力数据的占比将逐步提高，这给企业工作流程和我们的日常生活带来更高的效率和自动化。总之，数据在不断地增长，与此同时，数据真实性和及时性的重要性也在不断地提高。

我们在预测和分析数据驱动型世界时总结了如下经验：

- 随着数据变得对生活越来越重要、对企业越来越重要、更加实时化和更加移动化，管理和存储数据的实体必须开发出标准的方法，来提升可靠性、降低时延并提升安全性。这个过程可以从审核开始，但需要投资、连贯的战略以及一流的 IT 人才提供支持。
- 分析从事后走向实时甚至预测领域，分析在循证决策方面的使用情况逐步增加。这不仅涉及企业流程的数字化转型，而且还涉及企业的文化和组织结构。分析将成为一种竞争优势。
- 不可忽视安全和隐私方面的挑战。数据泄露可能迫使公司停业，有针对性的攻击可能阻止企业正常经营，黑客攻击可能损害商业机密。企业的业务、IT 和安全专业人员必须不断强调：安全不仅仅是一个需要纯技术解决方案的 IT 技术问题，还是一项要求各级员工全员参与的企业需求。
- 由于所有数字化活动都迁移到了 IP 网络，所以在物联网的推动或迫使下，业务领导和 IT 部门之间必须联合开展运作。而在过去，他们仅习惯于支持后台和财务部门，以及运行业务系统的实验室、手术室、工厂车间、电网、有线头端等。物联网是未来几十年业务改进的基础性技术支柱之一，因此从现在开始，相关数据的优化使用便已成为了企业成功的关键因素之一。领导和技术整合对于充分利用物联网技术至关重要，至少可以避免混乱。
- 将全球数据圈提升到全新水平的各种发展趋势将发挥出总体效应，使数字化转型成为引领企业未来十年成功发展的重要全方位手段。它还将提升对第三方的依赖，包括云提供商、软件公司和基础技术供应商。因此，供应商的选择将更多地被视为领导职能和合作职能，而不是采购职能。企业将有赖于这种职能。

“数据时代 2025”预测的 163ZB 全球数据圈只是一个开始，因为我们预计，这个世界将日益趋于网络化和数据驱动化。技术年代的一个十年能够并且很有可能会带来无法预知的进步、用例、业务和彻底改变我们生活的各种服务，它们全都有赖于所谓“数字化生命线”的数据。存储行业及其所有参与者将会发现，大量客户希望储存他们宝贵的比特，正是这些比特将帮助推动全球企业和个人生活的积极发展，并构成我们全球数据圈的一部分。

## IDC 总部

美国马萨诸塞州弗雷明汉市  
斯宾大街 5 号  
邮政编码 01701  
508.872.8200  
推特：@IDC  
idc-community.com  
www.idc.com

### 版权声明

本 IDC 研究文件作为 IDC 持续情报服务的一部分而发布，该项服务包含书面研究、分析师互动、电话说明会和会议。请登录 [www.idc.com](http://www.idc.com) 进一步了解 IDC 订阅和咨询服务。欲查看 IDC 全球分支机构的名单，请登录 [www.idc.com/offices](http://www.idc.com/offices)。请致电 IDC 热线 800.343.4952 转 7988 (或 +1.508.988.7988) 或通过 [sales@idc.com](mailto:sales@idc.com)，了解如何申请购买 IDC 服务，或者了解其他副本或 Web 权限的有关情况。

未经授权，禁止复制。  
保留所有权利。

## 关于 IDC

国际数据公司 (IDC) 是全球著名的信息技术、电信行业和消费科技咨询、顾问及活动服务专业提供商。IDC 帮助 IT 专业人士、企业高管和投资机构制定以事实为基础的技术采购决策和业务发展战略。IDC 在全球拥有 1100 多名分析师，他们具有全球化、区域性和本地化的专业视角，对 110 多个国家的技术发展趋势和业务营销机会进行深入分析。在 IDC 50 年的发展历程中，众多企业客户借助 IDC 的战略分析成功实现了自身的关键业务目标。IDC 是 IDG 旗下子公司，IDG 是全球领先的技术媒体、研究咨询及会展服务公司。