

云迁移预测

危中有机，云享未来

Duncan Stewart, Nobuo Okubo, Patrick Jehu, and Michael Liu

过去十年，云计算的增长已成大势所趋。2015年，云计算市场实现了三位数的年增长。尽管到2019年底，大型超大规模公共云提供商的年增长率已下降至“仅”31%，且随着行业走向成熟，预计2020至2021年这一增长率将继续（缓慢）下降，但云计算市场的增长仍然领跑许多其他行业。

2020年，受新冠疫情和全球经济衰退影响，多个领域已削减开支，云支出下降本在情理之中。然而，云市场仍然呈现出强大的韧性。从某些指标上看，2020年云市场增速基本与去年持平；按其他增长衡量指标，即便面对现代史上最严峻的经济收缩，2020年云市场增速仍然较2019年更快。原因或许在于：受新冠疫情、封锁措施和随时随地办公模式驱动，云需求出现增长。我们预测，为了节省资金、增强灵活性以及推动创新，企业将纷纷上云，因而2021年至2025年，云收入增长将与2019年持平或更高（即大于30%）。

“在两个月内，我们见证了本需两年才能实现的数字化转型。”

— 微软首席执行官Satya Nadella, 2020年4月¹

云增长概览

当然，云并非唯一可行的解决方案。从整个公司层面来看，几乎没有任何系统仅仅依赖本地部署、公共云或私有云中的一种。大多数系统部署都融合了公共云和私有云，它们彼此不同，却又相互关联，这种方法被称为混合云。混合云有多种形式，如私有云与公共云相结合，或公共云与本地资源相结合。²通过采用多重部署模式，混合云可为用户带来诸多益处。

显而易见的是，混合云已然成为新常态。根据2020年3月的一份报告，2022年全球逾90%的企业将依赖混合云。³同月的另一项调查发现，97%的信息技术经理计划在两个或多个云上分配工作量，以最大程度提升弹性，满足监管合规要求，并获取来自不同提供商的领先服务。⁴

然而，即便在当前经济形势下，云提供商的未来依然可期。虽然许多企业仍会保有一些本地资源，但已纷纷加速向云端迁移。种种有形和可衡量的指标也表明云迁移正加速推进，包括工作量、公共云提供商收入、半导体公司收入（生产用于云平台的芯片）以及全球电信网络云流量增长。

值得注意的是，许多云预测如同黑匣子，均基于无法复制的专有信息。⁵然而，诸如超大规模云收入、数据中心芯片收入和云流量等指标都是公开信息，任何人都可以复盘我们的工作并从中发现相同的趋势。

本地部署VS. 云 (工作量对比)

2020年4月，一项针对50位首席信息官的调查发现，受访者预计本地部署的工作量占比将从2019年的59%下降至2021年的35%，减少了41%。⁶此外，受访者预计在同一时间段内，公共云的工作量占比将从23%增加至38%，私有云和混合云的工作量占比分别为20%和7%。68%的首席信息官将“迁移至公共云和/或扩展私有云”列为2020年主要信息技术支出，较六个月前一项类似调查上升了20个百分点。

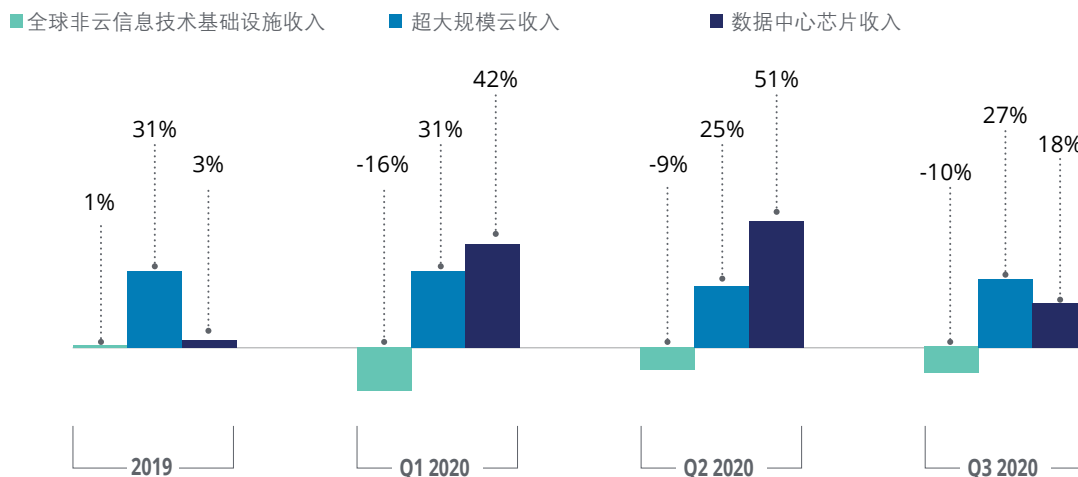
超大规模云收入概览

对于已披露细分收入的五大超大规模公共云提供商，2019年综合收入增长了31%，达到940亿美元。尽管2020年第一季度的技术支出普遍疲软，但收入较去年同比增长31%。第二季度收入同比增长25%，第三季度增长率小幅上升至27%，促使今年前九个月的增长率达到28%。2020年收入有望达到近1,250亿美元，到2021年将增至1,600亿美元以上。⁷尽管与2019年全年相比，2020年前9个月的增长率下降了4个百分点，但值得注意的是，以绝对美元计算，2020年整个超大规模云市场（不仅仅是五大市场）的增长率超过了2019年。2019年上半年市场规模较2018年上半年增长约400亿美元，而2020年上半年市场规模较2019年上半年增长约500亿美元。

图 1

云收入增长高于整体信息技术收入增长

收入同比增长率（百分比）



资料来源：德勤基于公开季度财务报表所作的分析。

数据中心半导体收入概览

截至2020年中，全球共有541个超大规模数据中心（2020年上半年新增26个），未来几年计划新增176个。⁸ 这些数据中心都需要芯片。虽然芯片支出和云收入并不完全关联，但从长远来看它们彼此依存，芯片收入增长往往是一个先行指标：在实现云收入之前，需率先购买芯片并安装于数据中心。⁹

对于已披露数据中心细分销售情况的三大半导体公司，2019年总收入仅增长3%，略低于300亿美元。在2020年第一季度，其收入较去年大幅增长42%。在第二季度，其收入同比增长51%。尽管第三季度的增长率下降至15%，但2020年前9个月增长率仍达到36%。2020年总收入有望超过350亿美元，到2021年或将突破400亿美元。

全球云流量

根据一份针对2020年2月1日至4月19日网络流量的深度检测报告，全球云流量在互联网总流量的占比从1.26%上升至1.83%，增长了45%。¹⁰ 同一时期，互联网总流量增长了38%，这意味着云流量（以每日绝对比特数量计）几乎增长了100%。

多维视角下的云增长

仅有约10家大型超大规模公共云提供商和芯片公司公布了详细的季度云收入。另有许多其他公司负责销售芯片、存储和互联解决方案，用于拓展云服务。尽管这些公司未提供详细的季度数据，它们对云市场的看法与已发布确切数据的公司相契合。例如，服务多个市场的存储供应商Micron在2020年第二季度表示，“由于居家办公、在线学习以及全球电子商务活动显著增加，大大刺激了云需求，我们的云DRAM销量季度环比大幅增长。”¹¹

投资者致力于加大云领域资金投入。截至2020年8月中旬，三大云交易所交易基金的总资产达到60亿美元，较年初增加了20多亿美元。在资产规模扩大的同时，业绩也得到了提升，三大云交易所交易基金截至10月30日的年均收益率为47%，而纳斯达克指数和标准普尔500指数的收益率则分别为22%和1%。¹²

投资者致力于加大云领域资金投入。截至2020年8月中旬，三大云交易所交易基金的总资产达到60亿美元，较年初增加了20多亿美元。

受新冠疫情影响，云需求扩大，从而推动了并购活动的增加。仅在2020年前4个月，已完成28宗数据中心收购交易，交易规模达75亿美元，超过了2019年全年的交易价值。¹³

注意事项

尽管云市场整体增长迅速，但云服务广泛涵盖诸多行业，其中多数行业已大幅削减开支。这意味着虽然总体增长强劲，但并不均衡。正如行业出版物《Siliconangle》所述：“由于大型云提供商规模庞大，它们所涉及的行业受到了疫情的严重冲击。因此，即使在这些公司，也出现了支出放缓的迹象。”¹⁴

未来前景展望

尽管2020年前9个月云市场增势强劲，但预计未来几个季度将在一定程度上放缓。然而，在两项因素的驱动下，云增长的放缓速度可能低于市场预期。首先，尽管不大可能出现2020年初的大规模封锁，但局部疫情的突然爆发以及更加本地化的封锁措施，仍将推动随时随地办公模式和云增长。其次，从长远的角度，许多员工和雇主认为疫情下被迫采取的随时随地办公模式行之有效。¹⁵例如，西门子允许员工在合理可行的情况下，每周两到三天采用随时随地办公模式；这项政策适用于43个国家及地区125个办事处的逾14万名员工。¹⁶随着企业进一步采用此类随时随地办公模式，对云服务的需求或将持续加大。

就行业格局而言，许多技术领域的市场集中度较高，一两家大公司几乎占据了所有的市场份额，云市场也不例外。2019年，前五大超大规模云提供商中，最大的两家占据了总收入的78%；同年，最大的芯片公司占据了数据中心半导体总收入的82%。¹⁷至少到目前为止，市场集中度并未因新冠疫情而进一步提高；事实上，虽然新冠疫情下云市场增速迅猛，但领先超大规模云提供商的市场份额却略有下降（下降了一至三个百分点）。基于过往的技术发展历程，从长期来看，随着经济增长恢复常态，市场集中度似乎有可能再次提高。在规模经济效益下，虽然未必出现“赢家通吃”的局面，但或许会迎来“赢家多吃”的市场。

全球超大规模云服务市场可能与中国走向“脱钩”，中国云提供商主要服务于中国市场，而世界其他地区则主要由美国超大规模公司提供服务。根据有限的的数据，截至2020年9月，中国超大规模云市场的增速似乎超过了世界其他地区，我们预计即便脱钩不会加剧，也将进一步延续。¹⁸

值得注意的是，除了受疫情影响之外，云迁移的需求由来已久。灵活消费模式，也被称为“一切（或万事）皆服务”，已然成为各行业企业日益重要的战略转型方向。虽然这一市场并不单单依赖于云，但云是一项重要驱动因素。截至2018年，“一切皆服务”的市场规模接近940亿美元，根据疫情前的一项预测，到2024年市场规模将超过3400亿美元，五年内复合年增长率达24%。¹⁹疫情可能会加速灵活消费模式的增长，但即使在后疫情时代，致力于云转型的行业参与者将更具财务可预见性，数据聚合的单位成本更低，并将进一步增强与客户的关系。已将业务转型为“一切皆服务”模式的公司在消费者和投资者中大获成功，不仅挑战了传统估值，还令那些固守传统商业模式（如永久许可和长期合同）的行业参与者压力倍增。²⁰

小结

为了推动持续增长，云提供商可采取以下几项举措。

俗话说，“欲戴王冠，必承其重。”高于预期的增长固然是好事，但为了保持增长，云提供商可能会面临更多的资金支出。2019年，超大规模资金支出（包括信息技术基础设施和实体基础设施支出）超过了1,200亿美元。²¹随着收入持续增长，超大规模资金支出或将继续以两位数增长，到2022年达到1,500亿美元。此外，不仅仅要加大资金支出上的投入。对于云提供商来说，人工智能应用程序和开发工具是吸引和维护企业客户的关键，因此也须加大投资或开展收购。²²

此外，随着基于云的企业工作流程占比从约三分之一增加至约三分之二，且上云速度较预期更快，关于隐私和安全的担忧亟待解决。例如，在疫情期间，医疗行业一直是最快上云的行业之一，因此可能更容易面临新的漏洞风险，尤其是在未有效实施云迁移的情况下。正如《Healthcare IT News》所述：“虽然云计算可以更好地优化医疗资源的配置，但也会带来重大风险，尤其是在云部署速度过快，信息安全人员难以及时开展有效尽职调查时。疫情过后，这一趋势仍将持续很长一段时间。”²³

超大规模云提供商的一个新兴发展领域是智能边缘。智能边缘将计算能力，特别是人工智能计算能力部署在终端用户附近（通常在50公里以内），而非集中化的数据中心。智能边缘并不会取代企业和超大规模云数据中心，而是提供一种跨网络任务分配方式，旨在提高及时性、互联性和安全性。²⁴在智能边缘模型中，以往上传至数据中心的大部分数据无须再进行传输，而超大规模云提供商需将这些数据回传至中心云，用于数据分析和人工智能培训，并确保数据访问不会中断。云提供商的另一个目标是开发垂直应用程序，由于延迟需求和其他因素，这些应用程序必须部署在边缘。超大规模云提供商可建立合作关系，以有效开展数据传输和应用程序开发。²⁵

对于云用户而言，在进行云迁移时应着重考虑以下因素：

云迁移策略。云迁移不只关乎于向云端迁移，还涉及持续的创新，确保借助云以降低成本，并创造新的机会。疫情之前，云迁移已经牵涉复杂。即便单个应用程序也可能关联多项业务流程，并对供应商、资产负债表和监管合规造成影响，而且不同的利益相关者可能有不同的动机和期望。一个简单的流程往往会导致目标冲突、依赖关系损坏以及成本超支。在后疫情时代，所有因素将更加错综复杂。在云迁移的过程中，关键任务是确保“在不中断业务的前提下，为市场带来颠覆式革新”。²⁶

云、安全和新冠疫情。如上所述，云用量的增加意味着更多的网络攻击事件，因此安全性比以往任何时候都更加重要，尤其在新冠疫情大幅推升云用量的背景下。根据2020年4月一项针对安全专业人士的调查，94%的受访者认为疫情会加剧网络威胁形势。近四分之一的受访者表示，与日俱增的威胁需要“引起重视且迫在眉睫”。仅15%的受访者认为网络威胁会在疫情过后恢复以往水平，而六分之五的受访者认为长远而言，将不断涌现新的威胁。²⁷

云成本和收益。疫情期间，多家企业迅速向云端迁移，其中部分企业的成本急剧上升。一些企业仅仅因为云用量增加，导致成本上涨20%至50%，不包括新应用程序或数据方面的成本。²⁸在云迁移的过程中，复制也会产生成本，企业需同时为云端、遗留系统以及两者间的数据同步支付费用。²⁹未来，企业应考虑成本规划（如通过折扣节省保留实例的费用），从而减少因仓促部署而导致的昂贵修复费用。成本管理系统也有助于费用管控。为了进行成本效益分析，企业可使用云价值计算器来评估当前状态与未来潜在机会之间的差距。这有助于优化基础设施，提高员工生产力，并增创业务价值。³⁰

全新价值机遇。向云端迁移不仅可促使企业稳健恢复，还有助于它们在疫情后全面反弹，不断提升韧性、促进业务连续性，继而开启新的发展并提供新的服务。未来，云将有益于推进协作、自动化、规模化、创新和敏捷性等。³¹以创新为例，在德勤2018年一项调查中，三分之二的受访者表示，云将助力他们全方位运用最新技术。³²另一项研究显示，93%的受访公司通过云来满足部分或全部人工智能需求，从而减少了在基础设施和专精知识上的投入。³³

在新冠疫情的推动下，企业纷纷上云，云市场有望迅速从危机中反弹，实现更加强劲的发展。随着云用量的增加，云提供商和生态系统其他参与者有机会不断发展壮大，而云用户则可做出新的尝试，依托云平台增创价值。云已经不仅仅是一种计算方法；在不久的将来，上云将成为各类业务的标准操作流程。

云与可持续发展

许多权威人士和组织，如世界经济论坛³⁴和经合组织³⁵，都呼吁在疫情之后启动“绿色复苏”。从这一角度而言，云计算有助于迈向更加可持续的未来。

2018年，云计算公司因能源消耗而屡遭诟病，甚至被冠以“耗能大户”的名号。然而，近来有文章指出，这些担忧有失公允。2010年至2018年间，云计算产能增长了600%，而其能源消耗仅增长了6%。³⁶

此外，云计算不仅产能高、耗能慢，而且能源用量通常比非云计算更低。平均而言，如将办公室应用程序部署于公共云，则每位员工每年二氧化碳排放约为6至7千克；如使用非虚拟化本地部署方案，则每位员工每年二氧化碳排放达到近30千克。³⁷在消费领域，截至2019年，观看一部两小时的Netflix电影，仅消耗Netflix及其发行合作伙伴8瓦时的能源，³⁸远远低于开车或乘坐公共交通去电影院所消耗的能源。

衡量云计算环境的关键指标是电能使用效率，云计算在这方面已经取得了长足的进展。大型数据中心的年均电能使用效率在13年内下降了36%，从2007年的2.5下降至2020年的1.59（数值越低越好）。超大规模数据中心则表现更佳，电能使用效率大约在1.1至1.2之间。³⁹ 鉴于此，近几年电能使用效率的提升已经放缓，超大规模云计算公司和芯片公司正通力合作，尽可能降低电能消耗。

云计算的能源需求呈现出两大趋势。第一大趋势显而易见，如果总体云用量翻番，那么在其他条件相同的情况下，能耗也会翻番。第二大趋势，人工智能在云计算中的占比越来越高。2012年中到2018年，用于前沿人工智能的云计算资源增长了30万倍。⁴⁰ 这在很大程度上归因于一种称为机器学习的人工智能，更确切地说，是机器学习的一个分支——深度学习。过去十年，深度学习在云计算中的应用越来越广泛，并拓展了许多新的功能。然而，为了不断优化深度学习，需依托更大的数据集进行训练，这意味着更多的计算，进而产生更大的能耗。来自加拿大的一位人工智能研究员表示：“令人担忧的是，机器学习算法通常会消耗越来越多的能源，使用更多的数据，并且需要更长的训练时间。”⁴¹

为了控制云计算不断增长的能源需求，可采取以下举措：

提高晶体管总体效率。随着时间的推移，晶体管变得越来越小。相较于15年前90纳米制程节点的半导体，如今10纳米制程节点的先进半导体在能效上提高了20至30倍。⁴² 未来，我们预计数据中心芯片将变得更加高效。

引进人工智能专用芯片。最初，所有人工智能计算都在通用中央处理器芯片（CPU）上进行。随着时间的推移，数据中心配置了更专业的人工智能专用芯片：图形处理器（GPU，用于训练和推理）、现场可编程门阵列（FPGA，主要用于推理），以及专用集成电路（ASIC，用于训练和推理）。⁴³ 虽然功耗效率视特定芯片和任务而有所差异，但大体而言，GPU用于训练时的效率是CPU的10至100倍，用于推理时的效率是CPU的1至10倍，FPGA用于推理时的效率是CPU的10至100倍，而ASIC用于训练和推理时的效率则是CPU的100-1000倍。⁴⁴ 随着人工智能专用芯片的广泛部署，数据中心的效率将不断提高。

优化人工智能专用芯片。用人工智能专用芯片取代CPU，可提高人工智能的功耗效率，但人工智能芯片技术目前仍处于初步发展阶段。例如，Nvidia推出了新一代人工智能GPU Tesla T4，比与其前身Tesla P4相比，在人工智能推理方面的能效提高了一倍。⁴⁵

将部分人工智能处理推向边缘。以往，由于智能手机、摄像头、传感器和机器人等边缘设备缺乏复杂人工智能所需的机载处理能力，几乎所有人工智能都在云端进行。相比之下，2024年的情形将大为不同，到2023年将有超过15亿台设备已配备专门的边缘人工智能芯片（称为神经处理器，NPU）。⁴⁶ 随着时间的推移，主要用于推理的NPU将具备更低的能耗：它们往往比数据中心的芯片更加高效，而且无须向云端发送大型数据文件，因而减少了能源消耗。与此同时，边缘人工智能芯片不仅相对高效，而且发展迅速。一家公司表示，现有边缘人工智能解决方案的效率较去年提高了25倍。⁴⁷

尾注

1. Jared Spataro, "2 years of digital transformation in 2 months," Microsoft, April 30, 2020. The cloud growth figures in the first two paragraphs are from Deloitte's analysis of publicly reported quarterly financial statements.
2. Wikipedia, "Cloud computing," accessed November 4, 2020.
3. International Data Corporation, "IDC expects 2021 to be the year of multi-cloud as global COVID-19 pandemic reaffirms critical need for business agility," March 31, 2020.
4. Business Wire, "Infrastructure and security challenges threaten multi-cloud and edge deployments, new survey from Volterra shows," March 9, 2020.
5. Wikipedia, "Replication crisis," accessed November 4, 2020.
6. Eric Jhonsa, "Public clouds are bright spot as IT outlays slow due to virus," The Street, April 3, 2020.
7. 德勤基于公开季度财务报表所作的分析。
8. Synergy Research Group, "Hyperscale data center count reaches 541 in mid-2020; another 176 in the pipeline," Globe Newswire, July 7, 2020.
9. 德勤基于公开季度财务报表所作的分析。
10. Sandvine, *COVID internet phenomena spotlight report*, accessed November 4, 2020. In the previous paragraph, the semiconductor growth figures are from Deloitte's analysis of publicly reported quarterly financial statements.
11. Farhan Ahmad, "FQ3 2020 financial results," Micron, June 29, 2020.
12. ETF Database, "Cloud computing ETF list," accessed November 4, 2020.
13. James Bourne, "2020 to be a 'bumper year' for data centre M&A—amid Covid-19—as value surpasses 2019 total," Cloud Tech, April 24, 2020.
14. Dave Vellante, "Cloud computing market remains strong, but it's not immune to COVID-19," SiliconANGLE, August 15, 2020.
15. Benedict Evans, "COVID and forced experiments," April 13, 2020.
16. SightsIn Plus, "Siemens to let staff 'Work From Anywhere' permanently," July 17, 2020.
17. 德勤基于公开季度财务报表所作的分析。
18. 德勤基于公开季度财务报表所作的分析。
19. IMARC Group, "Global anything-as-a-service market to reach US\$ 344.3 billion by 2024, organizations undergoing digital transformation facilitating growth," press release, accessed November 4, 2020.
20. Deloitte, "XaaS models: Our offerings," accessed November 4, 2020.
21. Synergy Research Group, "Hyperscale operator spending on data centers up 11% in 2019 despite only modest capex growth," March 24, 2020.
22. Stephanie Condon, "Google expands its Cloud AI portfolio," ZD Net, September 1, 2020.
23. Mike Miliard, "Hasty rush to cloud hosting during COVID-19 crisis could set stage for 'cyberpandemic'," *Healthcare IT News*, June 30, 2020.

24. Chris Arkenberg et al., *Unbundling the cloud with the intelligent edge: How edge computing, AI, and advanced connectivity are enabling enterprises to become more responsive to a fast-moving world*, Deloitte Insights, September 8, 2020.
25. Kelly Hill, "Google Cloud jumps into 5G and telco edge computing with a 'GMEC' vision," RCR Wireless News, March 5, 2020.
26. Deloitte, *Cloud migration*, 2019.
27. Blackhat, *Cyber threats in turbulent times*, accessed November 4, 2020.
28. David Linthicum, "COVID-19 leads to shocking cloud computing bills," InfoWorld, July 31, 2020.
29. Mike Chan, "How to calculate the true cost cloud migration," Network World, February 2, 2017.
30. Deloitte, "Cloud value calculator," accessed November 4, 2020.
31. Deloitte, *In times like these you learn to love the cloud*, accessed November 4, 2020.
32. Gillian Crossan et al., *Accelerating agility with XaaS: Many companies are using IT as-a-service to steer their way to competitive advantage*, Deloitte Insights, 2018.
33. Beena Ammanath, Susanne Hupfer, and David Jarvis, *Thriving in the era of pervasive AI: Deloitte's State of AI in the Enterprise*, 3rd edition, Deloitte Insights, 2020.
34. CB Bhattacharya, "How the great COVID-19 reset can help firms build a sustainable future," World Economic Forum, May 15, 2020.
35. OECD, "Building back better: A sustainable, resilient recovery after COVID-19," June 5, 2020.
36. Steve Lohr, "Cloud computing is not the energy hog that had been feared," *New York Times*, February 27, 2020.
37. NRDC, "Is cloud computing always greener?," issue brief, October 2012.
38. 2019年，Netflix在视频播放上的能耗为 451,000兆瓦时，Netflix用户全年浏览时长达到了1,120亿小时。
39. Rich Miller, "Data bytes: Cloud CapEx, PUE trends, HPC spending, servers on the edge," Data Center Frontier, August 3, 2020.
40. OpenAI, "AI and compute," May 16, 2018.
41. Will Knight, "AI can do great things—if it does not burn the planet," *Wired*, January 21, 2020.
42. Saif M. Khan Alexander Mann, *AI chips: What they are and why they matter*, Center for Security and Emerging Technology, p. 18.
43. Deloitte, *Hitting the accelerator: The next generation of machine-learning chips*, 2017.
44. Khan and Mann, *AI chips*, p. 23.
45. Nvidia, *Nvidia AI Inference platform*, 2018.
46. Paul Lee et al. *Technology, Media, and Telecommunications Predictions 2020*, Deloitte Insights, 2019.
47. Thomas Lorensen, "Arm Cortex-M55 and Ethos-U55 processors: Extending the performance of Arm's ML portfolio for endpoint devices," Arm Community, February 10, 2020.