

AI边缘计算技术白皮书

AI-ORIENTED EDGE COMPUTING TECHNICAL WHITE PAPER

2018-2019



|

联合发布

摘要

从 1950 年著名的图灵测试开始，人工智能（AI）发展已经经历了三大阶段。AI 产业正在从“感知智能”走向“认知智能”，人工智能作为工业制造 4.0 时代的核心驱动力，对产业结构、信息化水平和运营模式都将产生颠覆性的影响。

当前，AI 服务已经在工业制造、企业服务、生活消费等领域开始了广泛的应用。具备计算加速能力的 GPU、TPU、NPU 等 AI 芯片成为了数据中心、智能手机终端等计算部署不可或缺的一环。随着第五代移动通信技术（5G）的成熟和商业化推进，AI 的应用也逐步向智能制造、智慧交通、智慧工业等场景化趋势演进。

边缘计算作为算力架构优化的最重要的技术，不仅能满足未来 AI 技术在应用场景延伸、算力分布化部署等方面要求，同时也是 5G 网络、物联网网络发展的重要方向。

为了加深 AI 在边缘计算上技术融合、引导 AI 边缘计算技术发展、加速 AI 和 5G/ 物联网等相互赋能和边缘计算资源就绪。我们联合编写本白皮书，其主要内容包括：

- 定义 AI 边缘计算资源形态、技术体系和典型应用场景。
- 定义典型的 AI 边缘计算资源模型。
- 发布 AI 边缘计算关键技术。
- 提出 AI 边缘计算发展倡议。

01

边缘计算的总体思考

1.1 发展趋势和需求	1
1.2 定义及技术体系	3
1.3 边缘计算技术特性	5
1.4 边缘计算网络部署模式	6

02

AI 边缘计算典型应用场景

2.1 驱动力分析	8
2.2 边缘计算典型应用场景	9

03

AI 边缘计算关键性能模型

3.1 网络资源模型	12
3.2 计算资源模型	14

04

边缘计算关键平台功能

4.1 边缘计算 AI 加速	16
4.2 带宽和计算资源调度	17
4.3 安全隔离	18
4.4 移动性与边缘计算应用连续性	21
4.5 端 / 边 / 云协同	23
4.6 边缘计算 AI 计算框架	24
4.7 边缘计算微服务框架	24
4.8 能力开放	26

05

边缘计算产业拓展需求

5.1 标准化需求	28
5.2 OICT 商业模式	29
5.3 基础设施共享与平台合	29
5.4 AI X 5G 边缘计算赋能与融合	30

06

倡议和愿景

倡议和愿景	31
-------	----

07

参考文献

附录

联合撰写单位及编写人

联合撰写单位及编写人	33
------------	----

01

边缘计算的总体思考

1.1 发展趋势和需求

互联网服务市场纵深发展，带来了信息流量和计算需求的巨大的变化。根据多家机构预测 [1]，5G 将进一步刺激视频类富媒体流量的发展，移动视频流量每年增长 45%，到 2023 年占总体移动数据流量的 73%。全球互联网数据逐年提升，2020 年将达到 40ZB，其中 40% 流量都将由物联网产生，进而带来了数据分析和处理方面的极大需求。

边缘计算能够在靠近用户或数据源的位置提供网络、计算、存储服务，不仅能够实现流量的本地化处理，降低对传输网络和远端数据中心的流量冲击；而且能够提供低时延和高稳定的应用运行环境，有利于计算框架在终端和数据中心间的延展，有助于实现场景需求、算力分布和部署成本的最佳匹配。

边缘计算将满足消费者和产业互联网市场的如下需求。

一、边缘计算是更新消费者互联网体验的重要保障。

- **视频消费再升级：**一方面，消费者期待更高画质、缓冲零容忍的观看体验。从 1080P 到 4K/8K，从 30 帧到 60/120 帧，更高画质和流畅度不断刷新消费者的观看习惯，而画质每一个等级的提升都将带来流量带宽 3~4 倍的增长。另一方面，视频正朝着沉浸式和强交互服务方式发展。360 度全景、视频社交化推进视频流向多路和实时性方向演进。边缘计算能有效避免大流量对长途网络的资源抢占，确保从视频源到消费者的低时延、大带宽和高可靠。
- **场景化 AI 能力普及：**AI 能力已成为多数终端的默认配置。从简单图片处理到 AR

渲染，再到复杂的媒体编辑，AI 能力丰富了应用的交互手段，同时也拓宽消费者对于信息编辑和创造的想象空间。而不受时间、空间、和终端类型限制，为用户持续提供充足的 AI 算力则是智能化应用发展的趋势。边缘计算能够克服终端在功耗、存储容量等方面限制，兼具云端强大算力和本地端超低时延的优势，将为 AI 应用加速推广奠定更坚实的基础。

- **消费者多模交互：**物联网、5G 技术推进了万物互联，更多兼具计算及存储功能的设备联网，智能穿戴、网联车载娱乐、AR/VR 眼镜、无人机等连接设备的种类逐年增加。智能手机与消费者间交互时间将被多种终端分摊，消费向多模化方向发展。多模交互不仅带来了信息供给多元化，特定终端交互方式也对性能参数带来了极致要求，例如 VR 应用 20ms 以内的时延要求。边缘计算在满足多元应用性能要求的同时，也能通过分摊终端计算和存储能力的方式来实现终端设计轻量化的要求，进而降低终端成本、优化终端体验。

二、边缘计算是产业互联网发展的必要基础设施。

- **传统产业的数字化和网络化：**2019 年政府工作报告提出了新旧动能接续转换，包括传统产业升级、产业互联网构建、产业级的数字生态。其一，传统产业网络化加速了数据激增，例如智能交通在路边铺设高精雷达和多组摄像头，联动自动驾驶车辆来提高交通信息的精准抓取和高精模型建立，由此将带来 TB 级数据的产生。算法、算力的就近部署是这类产品规模化部署的基本要求之一。其二，产业数据的数字化及传播对信息安全提出更高挑战。私有化数据的不出网和安全保障需要边缘计算设施作为防护来达到部署的基本要求。
- **基础设施共享、产业中台：**为了最大化实现互联网技术对传统产业的赋能和效率提升，信息技术需要与行业诉求进行深度融合。通过数字化功能内嵌和产品软件化，实现智能化发展。边缘计算是产业互联网所依赖的重要资源，边缘计算平台带来的效率提升也将辐射到产业互联网的生产流程，带动数字化、流程化和平台服务发展。当前边缘计算平台化开源软件的持续升温，能够极大促进未来产业化平台效能提升。

- **数字产品向传统产业渗透和融合：**传统产业向互联网方向的发展需要维系数字产品与生产环境需求的内在联系。智慧城市、智慧交通、智慧农业、AI 安防等数字化产品也需要合理部署资源，才能服务于传统产业。边缘计算既满足了传统产业对部署安全等诸多方面的要求，通过平台化技术又能高效地满足数字化产品所需的部署环境。因此，边缘计算将促成数字世界与物理世界互动、融合，实现技术和市场的相互促进。

1.2 定义及技术体系

综合应用对于计算能力、时延和稳定性的要求，根据计算所执行位置和产品属性来分类，边缘计算可分为三类形态：

- **物边缘：**数据源自身具备计算能力，如智能手机、计算卡、智能摄像头等。由于在数据源本地处理计算请求，虽然响应时延低和稳定性能达到最优，但受限于功耗、物理资源等因素，一般仅完成简单计算任务。
- **移动边缘：**伴随 5G 网络发展而产生的新兴边缘计算资源，服务能配合 5G 终端、无线基站实现最佳边缘位置的布放，并结合 5G 能力开放可获得、定义无线带宽、网络切片、终端位置等增强功能。
- **云边缘：**在现有云服务中的 CDN 节点资源上，通过增加计算型服务器资源，在原有的流量加速服务上开展函数计算、AI 智能服务等功能。

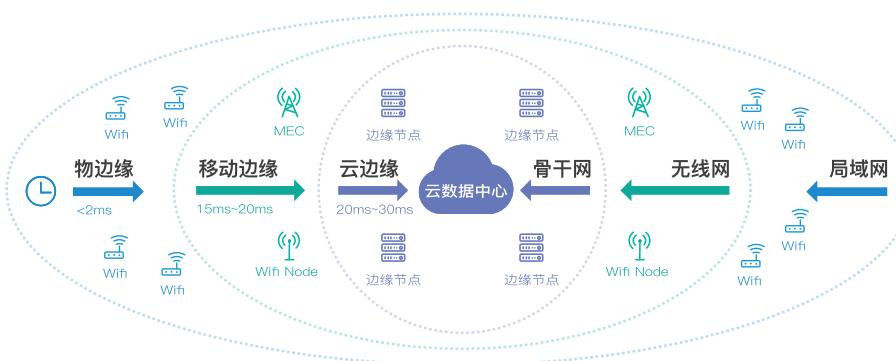


图1 边缘计算资源形态

以上三类边缘计算资源在场景设计、计算任务、产品属性和资源就绪程度上各不相同。在处理复杂 AI 计算任务上，需要能够集中各自资源优势，实现计算任务和相应资源匹配，才能取得总体最优的效果。



图2 边缘计算技术栈

为了达到这样的效果，边缘计算技术体系需要兼具物理资源属性调优、平台逻辑资源优化和高效计算任务框架，实现硬件和软件层最佳融合。如图 2 所示，技术体系分为物理资源及加速、IaaS 平台资源管理、PaaS 服务、AI 算法框架和应用服务五部分内容。

- **边缘计算物理资源：**物理资源分为智能终端、移动边缘计算站点和云边缘站点，根据不同资源程度来分配计算任务。三类资源通过接入侧或广域侧网络技术完成节点资源的连接。接入侧网络包括 4G/5G、固定接入、物联网或 LAN 技术，完成端计算到其他边缘站点资源的连接。广域侧通过 PoP 点完成数据中心云端到移动边缘 / 云边缘站点的连接，保障边缘与云端的网络质量。
- **物理资源加速：**完成了以上边缘节点本地计算、存储、I/O 优化和节点连接的加速优化。端节点通过 FPGA 等硬件加速技术为 AI 推理计算提供良好的性能环境。移动边缘和云边缘站点在适应功耗要求下通过 AI 芯片、存储优化、高速 I/O 框架实现加速和效果提升。此外，边缘节点间连接加速可通过 5G 切片、QoS 调优或 SDN-WAN 等完成节点间连接的加速功能。

- **平台资源管理：** 资源管理实现对 CPU/GPU、存储和网络的虚拟化、容器化和池化功能，满足资源的弹性调度和集群管理要求。同时，平台将完成租户隔离、安全保障、镜像仓库、日志存储等资源层面服务的供给。
- **PaaS 服务：** 面向服务提供三个阶段功能，一是应用设计及开发阶段的微服务化，通过微服务框架完成微服务模块间依赖关系绘制和模块到边缘计算资源的服务编排。二是提供服务的运行态环境和通信框架，例如业务发现、MQTT、RPC 等为服务运行提供框架性的功能支持。三是对于服务 CI/CD 和运行状态监控等管理面服务支持。
- **AI 算法框架：** 从延迟、内存占用量和能效等方面，进行边缘计算节点上 AI 推理加速和多节点间 AI 训练算法的联动，完成轻量级、低时延、高效的 AI 计算框架。边缘设备需要执行越来越多的智能任务，例如智能驾驶需要街道目标检测识别，语音助手需要自然语言的理解。针对实时信息输入，AI 算法需要执行预测处理、及时响应输入。此外，针对信息安全、数据不出网等要求，边缘节点需要完成数据安全预处理，边缘 - 云数据中心协同才能开展完整的 AI 模型训练。以上操作配合 CV、NLP 等 AI 算法模型库和强化学习、迁移学习等工具组件，形成完整的 AI 算法框架。
- **AI 应用：** 边缘节点应用对计算和流量带宽处理存在较强依赖。计算方面，应用需要 AI 算法框架完成人机交互、编解码 / 加解密等算法框架进行信息预处理、交通 / 医疗等建模算法构建专业领域信息框架，同时 PaaS 提供了“write once, run everywhere”的友好环境。流量方面，边缘节点需要数据源带宽低收敛比、低时延响应的物理资源环境，以满足数据传输和交互需求。

1.3 边缘计算技术特性

从物理特性和逻辑特性来总结边缘计算的技术特性。

物理上，边缘节点具有近源端、安全亲和、结构扁平和基础设施规格化特征。

- **低时延和流量本地化：**靠近数据源所带来的低时延和高带宽是边缘计算节点的本质优势，具备多接入（Multi-access）特性的边缘计算会将这种技术优势惠及到不同类型的数据源上。
- **安全亲和：**在工业、交通、医疗等行业应用上，对数据安全、网络安全、信息安全的保障是边缘计算开展的必要条件和服务场景，边缘节点资源安全布放是相关场景开展的基础。
- **计算扁平化：**边缘计算使智能化进一步普及，算力可随着边缘节点资源得以延伸，实现了计算的去中心化和无边界化。
- **基础设施规格化：**不同的应用场景，对基础设施的要求也不相同。为提高应用部署效率，需对基础设施部署统一规格。规格化的定义可从部署安装环境、I/O及加速部件可扩展性、高温/高湿环境适应性、故障管理和设备易维护等几个方面来定义。规格化定义有助于从设备供应、安装部署、运营维护和故障恢复等形成一系列产业建议，促进生产链成熟。

逻辑上，边缘节点具有场景化、软件友好和服务标准化特征。

- **场景化和智能化：**边缘计算由于所覆盖人群和区域的定向性，由此也衍生了服务场景化，例如服务于工业、园区、交通等边缘在服务开展上具有鲜明的行业特征，智能化也可以根据服务场景做定向算法优化和AI赋能。
- **软件友好：**边缘服务是云服务的延伸，是现场服务的扩展，是整体软件服务一环。通过边缘计算PaaS框架屏蔽底层硬件差异和资源平台差异，实现云-边-端服务软件统一部署。
- **服务标准化和模板化：**服务快速部署和弹性运维是开展规模化边缘计算服务的前提，这也要求对典型边缘计算场景和算力、带宽等资源进行标准化和模板化设计，实现快速复制和拓展的效果。

1.4 边缘计算网络部署模式

为了适应不同行业差异化诉求，边缘计算在行业中的落地会存在多种部署形式，每一个最终落地的边缘网络的组网架构与所覆盖的业务/网络需求、管理和运营模式以及商业模式相对应。下面归纳出了三大类典型的边缘计算网络部署模式。

- **企业私网：**此模式针对有很强的对网络自主可控的企业诉求，将完整网络数据面和控制面完全部署在企业私有环境中并交由企业自己管理。例如，在企业内部提供端到端的无线网络及其完整的控制面部署。专用网络与公共网络之间完全相互隔离，这样最大程度上保证了企业数据和网络环境的隔离和独立性，但是同时对企业自身的网络管理能力也提出了较高的要求。
- **园区专网：**在数据面上满足企业私有化部署要求，控制面复用公共部署资源。例如，5G部署时，可以将用户面网关（SGW-U/UPF）部署在临近企业现场位置，而5G控制面功能（AMF/SMF/PCF等）[2]仍使用网络运营商资源。又或者，云服务提供商为企业部署现场数据面处理功能，满足数据不出网等安全需求，而AI模型训练等离线控制功能仍使用云服务提供商资源。此模式既能确保企业的安全诉求，又能最大化复用公共运营商在资源管理和编排方面的能力，对技术轻量化运营的企业（如以园区为单位的一定数量的中小企业单位）不会产生过高的维护能力的要求。此外，该模式下公共服务运营商可以同时兼顾园区内边缘计算特定需求又能保证公网业务不受影响，这种模式很可能会成为边缘计算网络面向园区类市场的典型部署模式。
- **切片组网：**数据面和控制面均使用公共服务运营商的资源，通过划分专有的网络或资源隔离来实现专用化资源的供给。例如，5G采用网络切片技术提供资源隔离的逻辑网络。根据业务需求结合边缘计算来提供差异化的切片服务，构造逻辑上专用的网络，而物理资源则完全纳入公共运营商的统一调度。对于地域跨度大、且需要特定SLA保障的服务场景，切片组网是一个理想的解决方案。当然这种模式需要公共服务运营商对目标市场的判断和相应边缘计算网络资源的部署规划都提出了极高的要求，这也需要服务供需双方、设备提供商等产业各方共同探索来得出最佳实践。

02

AI 边缘计算典型应用场景

2.1 驱动力分析

边缘计算作为一种新型计算模型，驱动着计算架构由“端 – 云”升级为“端 – 边 – 云”。从业务供给的逻辑上，“边和云”主要提供业务所需的算力，“端”除了提供一定的算力外，还是业务的呈现者和消费者。灵活的边与分布式的端、集约式的云一起实现了算力的泛在化，进而为消费互联网升级和产业互联网的发展奠定了业务创新的基石。

手机大屏化、高清化和 AI 化驱动消费互联网业务升级。目前，6 英寸以上的大屏、1080P 的高清屏成为了市场上主流手机配置 [3]；未来随着“折叠屏”的发展，手机的屏幕尺寸和分辨率将会进一步增大，单手机的流量消费也将大幅提高。另外，随着人工智能技术的发展，AI 逐渐成为未来手机的杀手级功能。AI 智能手机可提供的 AI 体验包括：图像处理、视频处理、场景识别拍照、AR 特效、在线翻译等。

手机以外的场景化智能终端驱动新型消费互联网业务发展。随着语音识别、图像识别和计算机视觉技术的进步，催生了多样化的交互方式。基于新的交互方式，与特定场景深度融合的智能终端逐渐替代手机提供更好的业务体验，如：智能音箱、车载娱乐、AR 眼镜、VR 头盔等。其中，VR 被大家公认为下一代的革命性业务技术，极大的满足了人们的沉浸式体验需求；结合边缘计算的 VR 终端也将朝着无绳化、瘦终端、六自由度方向发展。未来，随着手势与姿态感知、全息技术等技术的进步，基于全息终端的“无屏”显示与交互将成为可能。

多样化智能工业端驱动传统产业由物理世界向数字世界演进，进而推动产业互联网的发展。产业互联网实质是将互联网与各个传统垂直行业深度融合，融合的前提是通过

智能工业端将“哑巴终端”改为“数字化终端”。每个垂直行业都有自己特定的智能终端，比如智能交通领域的路侧单元 RSU 以及 OBU、工业生产领域的传感器、控制器、执行器、安防领域的智能摄像头、电网领域中的智能分布式配电终端、分布式采集终端等。

智能端更新换代、AI 算力赋能丰富服务场景和服务市场向产业互联网转型都将推动边缘计算新场景和新应用的出现。

2.2 边缘计算典型应用场景

从边缘计算应用对信息消费形态来看，重点场景主要包括如下四个方面：

1) 边缘计算 + AI CV/NLP 类智能应用

边缘计算 +AI 加速了“Machine-to-Machine、Human-to-Machine”智能交互过程。基于 AI CV 模型的典型应用如人脸识别、图像检索、游戏和控制、安防监控、生物统计（指纹，虹膜，人脸匹配）和 AI NLP 的自然语言交互等目前都已有所应用，多以云 / 端解决方案为主。这种方式很容易受终端电池容量、散热、算力等因素限制，无法满足复杂性场景下 AI 应用对算力的要求。比如：移动终端集成“翻译”、“商品识别”、“AR”等功能在运行时，耗电达到 7~8W，是普通待机耗电（146mW 左右）的 50 倍左右。边缘计算可以较好的解决 AI 业务发展与终端能力之间的矛盾。比如：针对渲染类业务，在满足业务时延需求的前提下，通过云终端的形式将终端计算迁移至边缘云上，不仅较好的满足了业务的需求，而且降低了终端的成本。此外，在数据安全敏感的行业应用领域下，边缘计算在 AI 数据预处理环节也是不可缺少的。

2) 边缘计算 + 视频

视频一方面是网络流量的主要贡献者，另一方面也是对视频供给网络和平台的强依赖者。5G 将重构 4G 时代的视频业务，促使未来视频朝着超高清（4K/8K）、多维度（3D）、

沉浸感（VR）、交互性发展。细究这些业务特征，不难发现，大带宽、低时延、实时计算是对未来视频供给网络和平台的基本能力诉求。基于边缘计算的视频供给网络和平台可较好的满足上述需求。以 8K 高清直播为例，可运营级带宽需求为单路 120Mbps，端到端时延为 500ms 以内，通过 AI 技术对用户特性进行分析，预测其移动模式和内容请求的概率，在边缘计算节点有针对性地预先缓存内容，使更多业务请求在本地满足。

3) 边缘计算 + 云计算

边缘计算是云计算服务的延伸，云计算是边缘计算服务的基础。伴随 5G、物联网等典型边缘计算网络场景的推进，“云 + 网”服务更加深入人心。边 – 云协同促进了云网融合，能够完成 5G、物联网 LAN 与云端应用结合和快速部署，实现“边到云”能力拓展和“云到边”的场景赋能。通过在边缘计算节点部署混合云接入网关，可以帮助 5G 或物联网赋能的企业网快速上云，一方面可满足数据不出企业的安全需求，另一方面也可实现云端 AI 模型、集群应用到边缘的平滑过渡。

4) 边缘计算 + 智能交通等行业 / 企业专网服务

工业、交通、医疗、电力等行业服务是边缘计算应用的重点领域。由于行业应用所涉及流程和服务的专业性，边缘计算服务模型需要高度私有化定制。如下以智能交通、工业制造为例来阐述对行业 / 企业专网服务的需求。

智能交通正在逐步向“智能的路、聪明的车”方向发展。智能的路涉及到通过部署摄像头、高精雷达等路况感知设备对道路进行数字化改造。聪明的车通过与路侧感知系统交互，获取道路上车辆、行人等实时信息，再将路感信息与车感信息融合处理对周边环境进行全方位感知，最终实现完整的自动驾驶功能。上述“车路协同”对网络的时延要求为 10ms 以内，以高速公路为例，10KM 路段产生 1Gbps 的持续上行流量，与传统下行流量为主的服务模型有着很大的差别。这类场景下，在边缘计算服务模型部署、模型安全性保障和物理带宽资源上都需要作出针对性考虑。

伴随着新一代信息通信技术的发展和融合，制造业正迈向体系重构、动力变革、范式

迁移的新阶段，万物互联、大数据、工业互联网平台、智能化升级成为制造业新体系重构的主要发展方向。在制造领域，海量设备互联、海量数据采集分析、异构现场级设备通讯、超低时延工业控制等需求，需要采用边缘计算才能满足。同时，工业数据由于涉及到大量企业运营核心机密信息，在采集和处理过程中，需要进行比较高等级的安全保障。综合以上需求，在工业制造领域以企业专网方式提供设备连接基础、本地数据采集和预处理分析、数据和应用在边缘云上进行存储和处理，构建边缘计算 + 云网融合的一体化新型架构体系，可有效支撑制造行业的数字化、智能化转型。

03

AI 边缘计算关键性能模型

3.1 网络资源模型

数据中心下资源布局总体遵循 Hub&Spoke 模式，资源采用大区式集中，流量加速通过 CDN 节点完成省级布局。相比边缘计算，加速节点部署粒度稀疏，业务模型以通用流量调度为主、计算型服务为辅，因此当服务以计算型为主的行业应用或服务粒度较细的密集型应用时，利旧改造 CDN 节点会出现原有资源模型和边缘计算服务模型不匹配的问题。此外，随着光纤宽带等通信等通信基础设施的全面普及，加上 5G 技术成熟并商用发展，用户最后一公里接入和相应驻地网络的带宽将进一步得到大幅提升。更细资源粒度、调度策略和多层次（例如，园区级、驻地网级、区县级、地市级等）的资源布局将更加适合未来计算应用 AI 化、场景化的需求。

多层次的网络资源划分可以针对性满足特定场景对算力和网络需求，但也同时增加了资源投入和应用运维的复杂性。因此更需要面向典型应用的边缘计算资源块切割和标准化封装，不仅使得资源投入具有指向性，避免资源投入过剩风险，而且也便于应用运营时对资源着色和场景识别，做出有针对性的资源调度和供给优化。

根据 IMT-2020《5G 承载需求》[4] 中各级别资源分布及量化定义，以下通过典型应用场景、服务覆盖面积、服务人群、时延为例来定义边缘计算多层次的资源模型，如表 1。在实际生产部署中，划分边缘计算资源块标准可根据实际需求来定。总体原则是结合资源供给可行性和确定性，做到边缘计算资源的按需部署和应用运营针对性优化。

边缘计算层定义	带宽资源	覆盖人群	覆盖面积	延时要求	典型场景
驻地网级别	10Gbps~100Gbps	2~3万	2~10KM ²	移网:<15ms 固网:<2ms	智能交通、医疗、场馆、园区热点等
县区级别	100Gbps~1Tbps	<100万	<1000KM ²	移网:<20ms 固网:<10ms	UHD/VR应用区域性普及
地市级级别	≥1Tbps	≈1000万	>1000KM ²	移网:<25ms 固网:<15ms	广域覆盖要求的计算服务、VCDN等

表1 边缘计算资源块规划示例

案例 I – 智慧交通: 智慧交通业务场景要求极低时延及超高可靠性，同时对带宽要求也非常高（摄像头数据回传），适用于在驻地网级进行部署。摄像头带宽要求为 6–8Mbps，高精雷达带宽需求约为 200Mbps，按照单个路口部署 8 个摄像头和 1 部高精雷达测算，单个路口的带宽总需求约为 250Mbps。路口距离以 500 米计算，则每平方公里平均有 4 个路口，带宽总需求为 1Gbps。在驻地网级部署边缘计算节点，可以满足几平方公里至几十平方公里智慧交通业务的网络和计算需求。

案例 II – 居民园区或场馆热点覆盖: 场馆热点覆盖等场景要求较低时延及较大带宽（百 Gbps 级），该场景适用于在驻地网级网格进行部署。例如，现场观众通过高清视频从多个角度与场外观众互动分享体育赛事，场外观众可以从他们希望的任何视点观看比赛。该场景假定情况如下：体育场内平均用户吞吐量为 2Mbps，体育场座位宽 1 米，深 0.5 米，平均每 20 人有一位移动用户在通过视频为其他场外观众来分享比赛，体育场的用户密度为 $(0.0005 \text{ km} \times 0.001 \text{ km}) / \text{每用户}$ 。因此，当体育场内有 1000 个用户时，总计带宽需求为 $2\text{Mbps} \times 1000 \text{ 用户} / (0.0005 \text{ km} \times 0.001 \text{ km} \times 1000) \times 1/20 \times 1/1024 = 195.3125\text{Gbps}/\text{km}^2$ 。在驻地网级部署边缘计算节点，将可以满足几平方公里范围内场馆热点覆盖的业务需求。

案例 III – UHD/VR 应用区域普及：超高清视频（UHD）和虚拟现实服务（VR）需要更高带宽的支持，应用普及将可能从特定区县开始，逐步向地市和全国推广。根据测算，运营级的 4K 视频流为 60 帧 /s，最小带宽要求为 45Mbps；8K 视频为 150Mbps。运营级 VR 4K 满足强交互要求需要有 90 帧 /s 速率和 120°视场角，最小带宽要求为 80Mbps，VR 8K 最小带宽要求为 260Mbps。在区县一级边缘计算节点除完成 VR 渲染、视频编码等计算功能外，在百 Gbps 或近 Tbps 的网络环境下，可实现 4K/8K 视频 5–10% 同时在线的应用服务，或 4K/8K VR 应用 2–5% 同时在线服务，能够满足 UHD/VR 从小众类应用到普及类应用的过渡要求。

3.2 计算资源模型

在WWW时代，随着网页邮箱、搜索、电商、社交等应用的发展，互联网服务呈现出向云端转移的趋势，越来越多的计算和存储需求从PC客户端转向大型数据中心。数据中心可以通过共享降低成本，另外，搜索等应用也需要更加强大的计算能力，数据中心的规模越来越大，服务器、网络、机电等硬件设备需要协同工作，操作系统、数据库、中间件等软件需要与硬件协同工作，这样的数据中心称为仓储式数据中心（WSC，Warehouse-Scale Computer）[5]。

移动互联网时代，信息不再以单向下发为主，用户生成的内容越来越多，但是用户的终端设备如手机、平板等的存储、计算能力有限，大量的信息需要上传到数据中心进行存储和处理，所以进一步强化和扩大了WSC的规模。

物联网的快速发展让我们进入了后云时代，数以万亿的各种设备将会联网，摄像头、传感器等产生海量的数据，如果直接将物联网产生的数据传输给WSC，将会产生巨大的网络负载，网路可能造成拥堵，并且会有一定的数据处理延时。另外，时间敏感类型的应用要求极快的响应时间，数据敏感类型的应用则要求通信数据的私密性等等。所以，当前的云（WSC）端（手机等）协同的计算模式，不能满足物联网时代海量数据、快速响应、安全私密等的要求，边缘计算应运而生。在边缘节点处理这些数据将

降低响应时间、减轻网络负载、保证用户数据的私密性。终端-边缘节点-云数据中心(D-E-C)计算架构用以满足物联网时代的要求，如下表。

	Device	Edge	Cloud
算力	1–10 TOPS	10–100 TOPS	100+ TOPS
功耗	0.1–10 W	10–100 W	100+ W
延时	10–100 ms	ms~s	ms~s

表2 D-E-C 算力分布

04

边缘计算关键平台功能

4.1 边缘计算AI加速

在 AI 领域，专用芯片兴起，基于传统的 CMOS 技术将遇到制程和工艺材料的限制，学术界开始研究“Near data computing”，“processing in memory”，模拟计算，光子计算，量子计算等更前沿的技术。从产业界来看，当前 AI 芯片仍然以 Nvidia 为主，但行业已经有众多 AI 芯片公司在进入这一领域，并将在 19 年逐步有产品落地商用，如 Intel、Cerebras、Graphcore、Habana 等，20 年将有更多产品落地如寒武纪、华为等。随着边缘计算的逐步成熟，势必催生出更适宜边缘计算场景的芯片架构。因此训练和推理框架对下需要兼容并适配更为多样的硬件架构，对上要支持边缘计算典型的应用场景。由于边缘计算的特点，其硬件的异构化程度会显著高于传统数据中心，对现有计算框架会带来非常大的挑战。如何快速支持异构的计算芯片并保证计算的高效，非常值得产业内的研发力量持续投入。

从应用方面看，AI 在过去几年中在互联网应用、工业互联网、医学和生物学、媒体娱乐、安全以及自动驾驶等各个领域带来了突飞猛进的进展。硬件的进展是这次 AI 发展的基础推动力量。从供给侧来看，通用处理器性能增长终结，平均每年仅增长 3%；而 AI 计算需求量 2 年提升 10 倍。数据中心的 AI 推理与训练芯片，是各大互联网公司、传统芯片公司以及创业公司聚焦的战场。

业界正在从通用转到专用，使用特定领域架构（DSA）解决算力及功耗问题。下图总结了各种计算架构的不同：

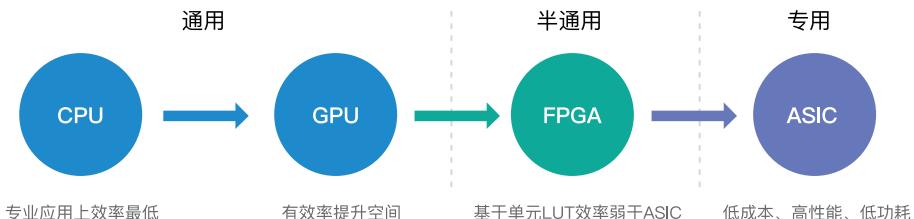


图3 计算架构加速

对于边缘 AI 计算总体来说，核心诉求是高性能、低成本、高灵活性。其技术发展趋势可以总结为以下几点：

- 可编程性，通用性
- 伸缩性，同一个架构支持不同场景
- 低功耗，适应更多边缘场景的环境和电力要求
- 软硬件深度结合，更高能量效率 (Tops/Watt)
- 高效的分布式互联和协作计算能力

4.2 带宽和计算资源调度

相比CDN，边缘计算调度不仅需要考虑带宽情况，而且也将涉及CPU、内存、GPU、存储、网络等各种算力资源的调度。

边缘计算的调度是包含多个层次的，既有节点内资源的调度，比如不同业务对资源使用的优先级，如何最优使用的同时又保证核心重要业务不受影响，即通过智能调度，边缘计算平台具备将业务随时迁入和迁出的能力；也包括节点粒度的资源调度，比如怎么就近接入，什么时候使用移动边缘，什么时候使用云边缘，亦或是把请求发回云中心进行计算，需要从多层级集群的角度进行统一的算力调度，因此边缘计算的调度要

求将所有边缘资源纳管，统一调度。

同时边缘计算的调度又可以从算法模型上，结合Device–Edge–Cloud统筹考虑。对于通用的计算模型，搭建统一的计算框架，结合实际的场景，确定将什么样的计算模型调度到什么类型的边缘算力平台上，以达到最优的效果。

对于有状态的计算，状态的迁移也和调度密不可分，比如在发生网络切换或者位置移动时，如何通过调度保持状态的不间断，做到即使跨MEC的切换，状态仍然能保持一致或快速恢复。

边缘计算的调度也有特殊的地方，由于条件限制，边缘节点与中心云的连接不稳定，所以调度还要考虑边缘集群自治能力，即要做到将调度能力下沉到集群，节点甚至工厂设备级别，在脱离云端的管理之后，能做到一定程度的智能控制。当边缘计算节点再次和云端连接后，及时同步自己的状态到云端，确保云端和边缘调度的一致性。

用户最后一公里接入和对应的驻地网络的规模及量级大幅度提升，要求边缘计算在靠近用户侧有更精细的调度能力，因此整个调度框架需要在纵向划分多层，对应多层次的网络资源划分，不同的层级之间，可以使用相同的调度策略，但是不同的调度技术来进行资源调度分配。

所以边缘计算的调度需要综合考虑资源，成本，质量，时延，优先级等因素，既要全局统一，又要兼顾边缘自治，通过调度和计算框架的结合，智能的将不同的请求，调

4.3 安全隔离

边缘计算需要保证多业务和多租户之间的安全隔离，即包括网络层面的隔离和资源层面的隔离。

网络隔离需要选定合适的网络虚拟化方案，首先保证业务与业务间在边缘站点内互访

的隔离，以防高风险的业务的恶意攻击；其次，不同业务的网络带宽会出现竞争，需要做好业务的带宽限速，以防单一业务带宽占用过高导致其他业务不稳定；此外移动网络的网络切片能力也是一个网络资源隔离和管控的方案，不同网络切片所覆盖的业务在实现跨地域的专有网络性能保证的同时也达到了相互之间非常理想的安全隔离效果。最后，网络安全方面，需要考虑防御SYN Flood、ACK Flood等攻击，通过异常流量自动检测，自动牵引和清洗，确保单一业务的攻击不影响其他的业务。

相比于云计算中心，边缘站点的资源隔离考虑的因素更复杂。边缘机房受条件限制，规模较小，考虑到站点空间和供电要求，布放服务器数量将受限，因此越靠近数据源的计算资源越珍贵。按照传统方式，云数据中心一般采用标准的虚拟机进行隔离，由于机房规模大，虚拟机对资源的消耗影响有限，但是在边缘机房，这种影响就比较明显。在私有部署环境下，使用更加轻量级的容器化方式，资源占用小、性能高，且能够提供基于Cgroup和Namespace的隔离，一般能满足大部分的自用场景。

但容器由于是共享内核，容器间并非完全隔离，对于外部多租户的情况下，安全风险较明显，而且容器内逃逸目前仍是业内需要攻克的难题，存在如下风险：

- 多租户之间资源的完全隔离，避免一些内存泄露，崩溃，对系统组件的错误调用等可能导致机器级别全局的风险；
- 高风险的进程隔离，防止有恶意代码，从容器内逃逸，进而对宿主机、边缘集群等进行攻击。
- 对内核有特殊要求的业务，容器共享内核，则无法按照客户的要求实现内核的定制化。
- 可信执行环境的构建，采用容器构建可信执行环境现在还没有现成的方案。

在云数据中心下通常使用KVM等内核级技术进行隔离，但考虑到边缘资源的有限性，直接使用KVM的场景有限。更加轻量的虚拟机和安全容器逐渐受到关注，其目的希望做到容器级的轻量快速，虚拟机级别的安全隔离。业内也有一些比较新的开源方案，虽然还不是很成熟，但已开始了快速迭代，比如Kata Container、Firecracker等，通过裁剪不必要的启动组件，确保VM快速启动，同时结合容器的特性，可对外提供完整的服务。

边缘计算承载的场景多样，既需要标准虚拟机和轻量级虚拟机的安全隔离满足对外客户的服务，也需要常规容器、安全容器等基本的隔离，因此需要统一的部署和调度。由于边缘计算平台多是采用Kubernetes，业内也在此基础上提出多种VM编排开源方案，例如Kubevirt和Virtlet。对于Kata Container，由于满足OCI标准，可以通过Containerd或者CRI-O进行集成，并且Kata Container已经尝试兼容Firecracker来实现其虚拟机的能力。

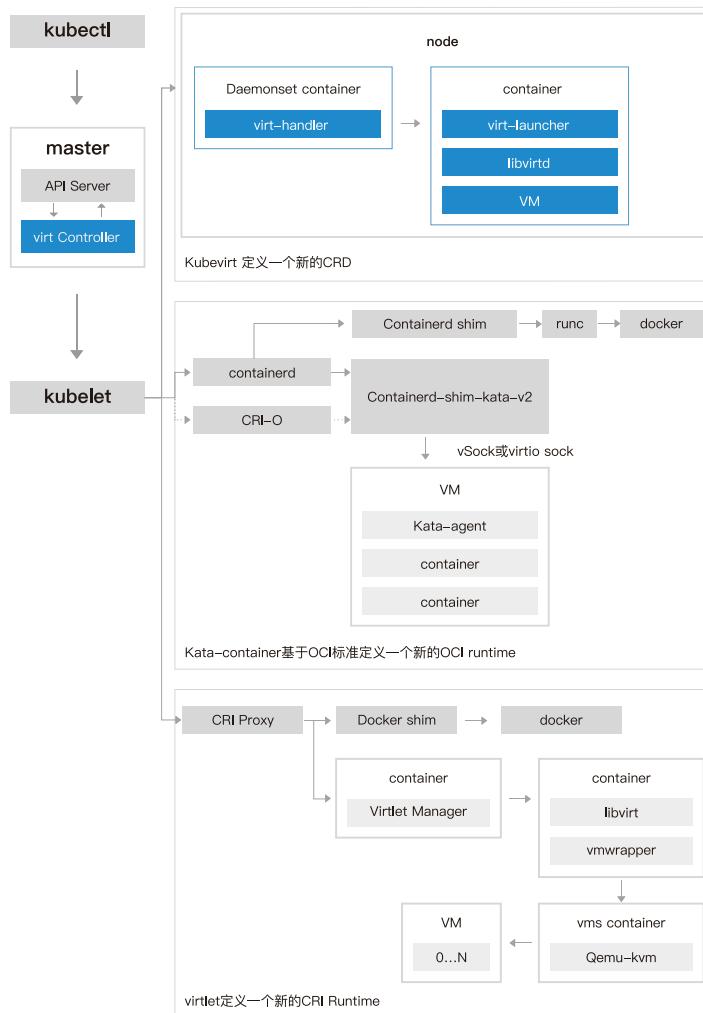


图4 VM和容器的混合调度的开源方案参考

相信随着边缘计算的多元化隔离需求的推动，轻量级虚拟机将很快成为一种主流的边缘计算底层虚拟化技术。

4.4 移动性与边缘计算应用连续性

用户移动会导致终端数据到边缘计算应用的网络路径发生变化，并且有可能出现从边缘计算节点到非边缘计算节点，从非边缘计算节点到边缘计算节点，或者从边缘计算节点A到边缘计算节点B的切换（受边缘计算节点的覆盖区域限制）。如何保证用户会话和业务的连续性，保障用户一致的业务体现，是边缘计算落地的重要问题。

从移动网络角度，对各种切换场景有相应的移动性管理机制（如基站间切换、边缘UPF间切换、集中UPF间切换等），保证用户在发生移动性时，会有连续的网络接入服务。从边缘计算平台角度，也需要在一定程度上对应用发生迁移时的业务应用连续性提供保障。边缘计算最大的优势是降低时延，但边缘计算部署位置越低，越容易发生终端用户移动而导致的边缘计算节点间切换以及业务应用迁移的可能。所以需要最大程度降低切换带来的时延增加，并且对于有状态型应用，还需要对相应业务实例、应用状态信息等在节点间进行迁移和终结。

4.5 端/边/云协同

边缘站点根据其资源布放和服务场景不同又分为了多个层次，例如工厂或者园区的私有边缘、5G基站或汇聚点的移动边缘、CDN机房的云边缘等。多层次边缘的协同，结合设备和云端的处理，优势互补，共同完成AI的大部分计算场景。协同计算也将包含多个层面：资源协同、算力协同以及数据的协同。

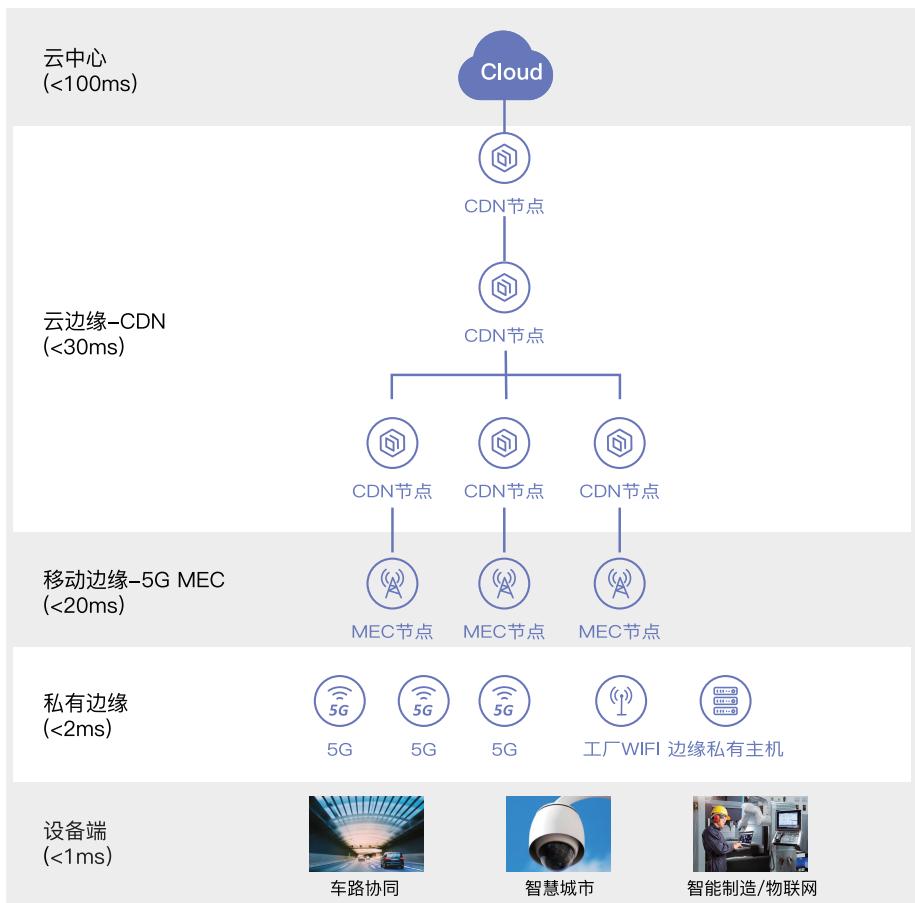


图5 多层次协同计算

首先是资源协同。算力资源在云-边-端间最佳分布为权重因子，实现包括带宽等资源在内的协同和调度。传统CDN中主要是以带宽最优使用为评价目标，比如CDN的多级回源，通过不同层级实现不同的回源策略，实现初步的资源协同，减少源站的带宽占用。而对于边缘计算，需要对算力资源有全局的判断，比如设备中的资源使用情况，边缘的资源使用情况以及中心云资源的情况，站在全局的角度，进行资源的合理分配，确保性能、成本、服务最优。

其次是算力的协同。通过合理的模型拆解，将不同的服务模型根据资源、成本、质量、时延等要求部放在合适的位置。通过完善的协同计算框架，确保各服务模型之间的协同处理。通过可靠的消息机制，确保状态的一致性。此外，结合产品的特性，进行综合设计，确保服务的体验流畅。以手机的视频搜索为例，将视频搜索的流程拆解为多个服务过程，比如可以分为物体分类、物种分类、精细识别等。结合产品的设计，将物体的分类放在手机端，只需要完成判断视频中的物体是属于哪一类，是植物，动物，还是人造物品。终端识别的结果和原始数据发给边缘端进一步的识别。边缘站点根据物体的分类进行物种的识别，比如对于动物，可以识别为猫科动物，做出识别之后再将结果和部分特征数据发给云端，进行精细的识别。云端根据提供的数据，结合云端完善的数据模型和知识体系，给出该猫科动物是老虎还是猫，是东北虎还是华南虎等。通过D-E-C的协同计算，渐进的识别出最后结果，既保证了用户的体验，又能确保合理的使用资源。

最后是数据的协同。这里的数据协同最主要指状态的协同。边缘的网络环境多种多样，设备的移动性将导致边缘接入环境经常变化，比如从某写字楼中走出来，可能就涉及终端的网络变化，从WiFi切换为4G/5G移动网、所依附的运营商网络变化等，网络的变化将导致边缘计算节点变化。对于有状态的计算服务，为了保证边缘服务的一致性，缓存的状态数据需要及时同步更新。边缘站点的各个层级、设备以及云端数据的协同将非常重要。通过边缘间状态的周期性同步，将数据恢复到上次同步后的状态，也可以从设备端快速恢复重建当前状态。另外边缘云基础设施所提供的虚拟机和容器实例，也涉及到实例的迁移，需要在数据协同方面做好支持。

4.6 边缘计算AI计算框架

边缘计算能够实现数据源上复杂计算任务的分担，同时也能优化云数据中心部分实时性要求较高的计算任务下沉。目前许多通用的AI学习算法框架，如TensorFlow、Caffe、PaddlePaddle等，通常更多地运行在云数据中心，无法直接应用于边缘设备。由于边缘计算和云数据中心所具备服务器容量和信息特征输入不同，因此AI计算框架设计上所对应的需求也较大不同。

在云数据中心算法执行框架更多地执行模型训练任务，他们接收大规模、批量化信息数据，关注于训练的迭代速度和收敛率等。而边缘设备更多地执行预测任务，输入是实时小规模数据，由于计算资源和存储资源相对受限，因此多数执行预测任务和少量具有实时性要求的训练任务，更加关注于预测算法执行时速度和训练计算时内存占用量和能效。

目前业界在边缘设备AI执行框架上也提出了针对性的设计方案，例如用于移动设备和嵌入式设备的轻量级解决方案TensorFlow Lite, Caffe2和PyTorch等。通过优化移动应用程序内核、预先激活和量化内核等方法来减少执行预测任务的延迟和内存占用量。此外，边缘计算在AI训练提速、安全信息预处理、边-云一体的AI算法上目前仍处在起步阶段，面向轻量级、高效、可扩展的边缘计算框架进行AI总体框架设计是实现边缘智能的重要步骤。

4.7 边缘计算微服务框架

微服务是一种软件架构风格，以专注于单一责任与功能的小型功能区块为基础，利用模组化的方式组合出复杂的大型应用程序。微服务将以全自动的方式部署，各功能区块使用与语言无关的 API 相互通信，同时服务会使用最小的规模的集中管理（例如 Docker）能力。



图6 边缘计算微服务框架

微服务化要求将复杂的问题进行拆解、解耦，使用标准的RPC或API通信，使得所有功能模块化开发，更容易理解和维护。每个微服务可以独立部署，单独测试上线，使得CI/CD成为可能。微服务上线后通过服务的注册和发现功能访问其他的微服务或其他微服务提供访问。由于微服务模块都不大，但数量较多，彼此之间的调用很频繁，因此调用链的追踪和分析对于问题定位特别重要。微服务架构要有完善的故障管理能力，包括但不限于限流、熔断、动态路由等，以确保服务故障的快速收敛。

微服务的小巧，快速开发，低耦合非常适用于边缘场景的快速迭代。目前第一代的成熟框架包括Dubbo、Spring Cloud等，但它们一定程度上仍与具体的实现语言和场景绑定，且只解决了Dev层面的问题，若想解决Ops的问题，还得和资源的调度框架结合使用。而Kubernetes作为容器资源编排的事实标准，并且基于微服务化的思想设计，容器作为最小工作单元，与开发语言无关，支持各种容器化应用的部署和管理，在边缘计算解决方案中已经成为主流，因此未来边缘计算的微服务架构应该会围绕Kubernetes展开。

服务管理的核心是流量的管理，服务网格作为服务间通信的基础设施，负责服务之间

的网络调用、限流、熔断和监控等。目前主流的服务网格开源软件有Linkerd、Envoy、Istio以及Conduit。考虑到和边缘计算的结合，Istio和Conduit对Kubernetes原生有较好的支持，它们的出现补足了Kubernetes在服务间服务通讯上的短板。

Kubernetes结合服务网格，在边缘计算领域可提供完整端到端的微服务能力。目前二者结合仍存在兼容和适配问题，需要随着边缘计算和微服务化推进，不断试错和完善。

4.8 能力开放

据Chinapex的预测，2018年API经济的全球市场规模预计将达到2.2万亿美元，在未来2–3年中，预计开展API项目的企业数量同比增速有望保持在100%以上。在中国，API经济规模将超过2万亿，能力开放市场潜力大，客户需求旺盛。

边缘计算通过API封装来为用户提供IT和云计算能力，主要包括本地分流能力、室内定位能力、TCP加速能力以及视频业务调度等关键的技术能力。

- **本地分流能力：**通信网络中数据统一在核心网出口。本地分流能力可以满足用户通过MEC平台直接访问本地网络，无需经过核心网，直接由MEC平台分流至本地网络，满足用户业务本地化处理能力。
- **定位服务能力：**满足室内导航、智能停车、智能办公等业务需求，通过室内基站收集用户测量信息并计算用户位置，MEC平台可将室内定位信息通过安全通道向签约可信的应用方提供，以此提升用户体验。
- **TCP加速能力：**通过MEC平台提供接入网络无线层信息，经过正向反馈来降低TCP拥塞率。TCP加速也可对用户内容的分辨率和视频质量做吞吐率引导，提升应用质量同时实现链路利用率的提高。

- **视频业务调度能力：**通过本地缓存、跨层视频优化、用户感知来实现对不同需求客户的区分和服务等级的定义，实现边缘计算平台的资源分配和业务质量保障，减少用户观看视频的卡顿，提升用户视频业务体验。

在基于SDN（软件定义网络）、NFV（网络功能的虚拟化）的云边端模式下，公共服务运营商可以提供不同粒度的网络功能，实现网络功能的灵活动态部署、易于调用，满足不同的应用场景对带宽、时延、移动性的多样化要求。边缘计算的开放能力通过能力开放架构来管理和实施，包括能力开放使能和能力开放管理。能力开放使能负责网络能力调用及相应的具体执行工作，向第三方应用提供北向接口，其南向接口用于对具体的网络功能实体进行访问控制。能力开放管理涉及网络能力开放的安全性及可靠性保障，既保证网络能力向可信的第三方应用开放，也确保第三方应用调用到可靠的网络能力。

05

边缘计算产业拓展需求

5.1 标准化需求

边缘计算的完整服务涉及服务需求方、资源供给方、设备提供方、平台提供方、服务提供方等多个环节和角色配合。标准化、模板化的服务定义和资源封装能便于不同角色间高效配合。

从技术构建到生产实践，标准化的需求主要来自技术形态上架构及关键功能的共识和生产形态上资源交付和服务的标准化封装。技术态和生产态下各环节和角色间的标准化是达成全链条无缝衔接和规模化部署的必要条件。表 3 为部分标准化诉求的列举。

	技术态标准化	生产态标准化
资源供给方	边缘计算层次定义、架构规范和接口定义	带宽资源、计算容量和SLO定义等
设备提供方	功能定义、规格定义和性能指标等	产品兼容性、设备能力可扩展性和产品形态演进计划
平台提供方	架构定义、功能框架和技术要求	平台调用接口格式、对接语义定义、分工界面划分
服务提供方	服务结构化定义	服务调用接口格式和服务

表3 不同形态下标准化诉求

5.2 OICT商业模式

智能化赋能领域正从 IT 走向 ICT，并最终向 OICT 全域和深度融合发展。以大数据、机器学习、深度学习为代表的 AI 计算框架已经在语音识别、用户画像等普通消费互联网领域广泛应用。在交通、制造、工业、电力等领域应用方兴未艾，对 AI 模型和算法框架设计上也提出了更多的需求和挑战。

OT (Operation Technology) 领域具有技术体系庞杂、专业性等特征，相比 ICT 标准化和开放性来讲，在成熟技术体系形成和多方协作方面存在更大的技术挑战。在这种趋势下，边缘计算是 OICT 产业场景服务最有利的资源支撑，而数字化和智能化是 OICT 转型的基础和目标。

通过在边缘计算上 AI 智能布局，满足生产、销售和服务的端到端行业智能需求，同时形成 OICT 最佳合作实践和成熟商业模式，成为多产业领域协作和商业化推进的重要手段。

5.3 基础设施共享与平台合作

从“端 – 云数据中心”计算到“端 – 边 – 云”AI 计算是对基础设施资源布局、智能端接入网络和应用计算模型的一次升级、甚至是颠覆。边缘计算所依赖的优质边缘站点以及网络、计算资源需要大规模投入和重资产建设。基础设施的共享是边缘计算产业发展的必由之路。

基础设施共享发生在两个层面：

- **一是资源层面的共享。**伴随 5G 产业对用户最后一公里网络的更新换代以及网络扁平化进程，边缘计算可获得更加优质的接入环境和接入层站点资源。此外，边缘计算也可随着行业智能化进程，在私有化站点部署下，协同行业场景开展边缘计算服务。

- **二是平台层面的共享。**结合资源供给方布局和应用需求，借力资源供给平台所提供的服务，开展应用平台服务与资源平台服务间的对接与合作，探索双赢的商务模式和加速边缘计算应用发展的商业化道路。

综合以上边缘计算发展诉求和路径，跨产业的资源共享和平台合作是不可或缺的。

5.4 AI + 5G边缘计算赋能与融合

源于 IT 的人工智能技术和源于 CT 的 5G 技术在边缘计算上达到了融合和统一。边缘计算是未来应用赖以开展的最重要基础设施。在此基础上，5G 作为基础网络技术将聚合来自移动互联网和产业互联网的诸多场景和诉求。AI 作为基础计算能力需要与丰富场景的结合才能完成从技术到商用化和产品化的蜕变。

5G 和 AI 结合既满足了 5G 对场景赋能的要求，又解决了 AI 寻求不同场景下产品化的要求，天然具有技术和市场优势互补的作用。

06

倡议和愿景

AI、5G、边缘计算所产生的聚变反应必将超乎想象，必将深刻影响消费互联网、制造、零售、教育等各行各业的发展模式。作为“AI 边缘计算行动计划”发起者，我们将促进 AI 与 5G 边缘计算深度融合发展生态，推动 AI 边缘计算技术和服务平台共享，包括如下三个方面的倡议：

- **共建边缘计算 AI 能力**

面向行业应用、互联网服务，共同定义边缘计算的 AI 能力。通过打造 AI 服务、搭建 AI 能力平台，为各行业发展提供智能化工具和服务手段。

- **共创边缘计算平台**

通过平台服务，为上层应用提供安全可靠、灵活部署、高效易用的运行环境，为 OICT 应用提供强大的算力、网络和存储资源支持。

- **共塑边缘计算品牌服务**

做好互联网行业服务升级和垂直行业数字转型的支撑，挖掘行业需求，进行跨界合作和场景创新，推动高价值、可规模发展的应用创新。

07

参考文献

[1] 爱立信移动市场报告

<https://www.ericsson.com/en/mobility-report>

[2] 3GPP, 23.501, System Architecture for the 5G System.

[3] IDC: 智能时代新竞争——2019 年中国智能终端市场十大预测

https://www.idc.com/url.do?url=/includes/pdf_download.jsp?containerId=prCHC44585318&position=77

[4] IMT-2020, 5G 承载需求白皮书

<http://www.imt-2020.org.cn/zh/documents/1>

[5] The Datacenter as a Computer, Designing Warehouse-Scale Machines,
Third Edition

附录

联合撰写单位及编写人

撰写单位



撰写人

百度：陈刚、孔德超、赵杰、庄熠、董勋、张贺纯、谷金龙、许振华

中国联通：韦广林、吕华章、张伟强

中国电信：赵龙刚，柏楠，徐丹，白燕南

中国移动：耿亮、宋雪飞、张晓秋

英特尔：许志豪、陈思仁、张骏

中兴通讯：王意军、朱堃

爱立信（中国）通信有限公司：甘泉、闫慧燕

浪潮：张磊

富士康工业互联网：吴仁杰

