

为基础设施提供加速引擎 助力塑造未来数据中心

基于英特尔® IPU 打造全新计算平台, 优化云平台管理能力更兼顾成本效益

目录

概述	1
背景及挑战	2
企业创新应用对云数据中心既有能力提出更高需求	2
实例多样化以及资源池化驱动向未来数据中心转型	2
英特尔® IPU 助力数据中心应对变革挑战	2
解决方案: 以英特尔® IPU , 构建云数据中心加速引擎	3
卸载基础设施能力与释放算力	5
对接英特尔® 至强® 可扩展处理器, 提升云数据中心算力 ...	5
助力百度智能云打造更高效可靠的 云服务能力	6
总结与展望	7

概述

为使云服务更高效、灵活地为各行业的业务创新与智能化转型提供支持, 特别是促进企业上云、边缘计算上云, 众多云服务提供商正致力于通过云数据中心架构革新和软件调优来消除基础设施的能力瓶颈, 使其更高效和智能, 从而为用户带来更强的服务性能、更高的稳定性与安全性和更优的管控效率。

针对这一需求, 英特尔推出了全新的基础设施处理设备——英特尔® 基础设施处理单元 (Infrastructure Processing Unit, IPU)。通过可编程阵列逻辑 (Field Programmable Gate Array, FPGA) 芯片加英特尔® 至强® D 处理器组成的硬件加速引擎或单独由专用集成电路 (Application Specific Integrated Circuit, ASIC) 芯片, 英特尔® IPU 能在软件加持下, 以可编程的方式将网络、存储和安全等工作负载从主机侧处理器上卸载并进行加速, 进而在充分释放云数据中心算力, 保持基础设施高性能、高利用率的同时, 实现更灵活地资源调配, 并使云服务提供商能更快地转向完全虚拟化存储和网络架构。另外, 英特尔® IPU 作为新基础设施能力加速引擎, 通过与英特尔® 至强® 可扩展处理器协同, 能充分满足用户在更多场景下的云服务需求。

作为合作伙伴, 百度智能云与英特尔在云数据中心未来建设方向上有着相同理念与愿景, 通过和技术层面的深度探讨和交流, 借力英特尔® IPU 平台参考设计, 自研百度太行 DPU 1.0 产品, 积极应对其在云服务实践中遇到的一系列挑战。新产品通过在裸金属服务器和云主机等场景中的部署应用, 被证明可成为其未来智能云服务的可靠技术基座。

背景及挑战

■ 企业创新应用对云数据中心既有能力提出更高需求

得益于技术与市场的高速发展,今天的云服务正为千行百业的生产经营、业务拓展和技术研发带来高效、弹性且充沛的算力支持,并助力更多用户基于云环境来灵活地部署和运行其企业级应用。随着 5G、人工智能 (Artificial Intelligence, AI)、边缘计算等前沿技术在更多行业获得落地应用,企业级云服务的发展也出现了许多新的变化与趋势,包括:

- **更呈分布式的云服务部署:** 边缘计算等技术的发展推动着云服务由传统的中心云逐渐向“云边端”协同演进,更多云服务被部署在贴近应用的边缘端;
- **更为广泛的云服务应用场景:** 云服务正成为更多行业的 IT 基础设施之一,在智能制造、智慧金融等新业态中担纲主力,成为各类企业级应用的坚实技术基座;
- **更复杂的工作负载需求:** 不同场景的工作负载对资源的需求不尽相同,有些负载需要更多算力,有些负载需要更大内存或者加速单元。如何满足不同工作负载的需求成为企业上云的考量之一。

■ 实例多样化以及资源池化驱动向未来数据中心转型

上述部署方案与应用场景的变化,也让云数据中心服务器在管理与应用上面临更多挑战。例如更多虚拟机 (Virtual Machine, VM) 的部署使虚拟机管理等管理任务变得更加复杂。同时,其能力输出也逐渐从传统单体式应用向微服务化发展,这也进一步提高了云数据中心管理任务的复杂程度,进而占用大量处理器资源。

既要对虚拟机、微服务,甚至裸金属服务器等实例开展高效管理,又要对相关的云数据中心的网络、存储等基础设施实现加速,其中的复杂性和资源开销不断增加,传统基础架构何以满

足未来数据中心要求就成为现实问题。换句话说,未来数据中心需要什么技术来实现转型?

随着微服务模型越来越多地用于云数据中心的应用,可以预测的是,未来数据中心可能具有以下的技术特性:

- 云原生应用的出现和快速增长推高了对专用基础设施的需求,并大幅提高云服务的敏捷性以及云数据中心的效率;
- 云服务的微服务化,推动了分布式异构计算环境的发展,而每个微服务都应在更适合的加速节点上运行;
- 微服务模型的广泛采用也催生数据中心编排系统,使各异构计算服务器之间实现微服务分发的自动化及管理;
- 使用微服务、虚拟机、容器以及容器编排的场景不断增加,推动了服务网格的开发。服务网格能够简化微服务到微服务的通信,并使其更加高效。现在,服务网格已经成为云原生堆栈的标准组成部分;

同时,由于不同工作负载或者用户不同时段对于资源需求的不同,通过资源池化可以更优的配置资源以达到弹性扩容的要求。

因此,未来数据中心的架构将更多地面向于服务不同的实例且日益资源池化的方向发展,以此来更好利用异构计算带来的加速处理能力。这一异构计算的核心将由传统的处理器平台,以及由英特尔推出的 IPU 等基础设施处理设备来组成。在基于这种架构的数据中心内,服务器与执行网络、存储等加速的 IPU 设备互联互通,不仅传统的服务器节点处理能力可在专用计算节点的支持下得到增强,同时存储、网络服务,以及云管理服务包括裸金属服务器和微服务也能获得加速。

■ 英特尔® IPU 助力数据中心应对变革挑战

随着多种实例服务以及资源池化等越来越普遍地用于云数据中心的应用、存储和网络,传统数据中心也在通过不断转型来

应对更高的需求和挑战，并由此向未来数据中心演进。在传统数据中心的架构中，通常借助传统的网卡（Network Interface Card, NIC）产品来负责执行物理层与数据链路层的数据流量处理，而更高层的功能则需要处理器等计算资源的参与。例如，在数据存储过程中，每一次 I/O 都需借助处理器多次上下文切换与内存拷贝，以“中断”模式在用户态与内核态进行数据交换。

而在提供云服务过程中，这一情况则更为突出。不仅 Open vSwitch (OVS) 等虚拟交换技术、RDMA over Converged Ethernet (RoCE) 等传输存储协议以及相应的数据安全技术被引入系统架构，同时也衍生出裸金属、容器云等多样化的云产品形态。这些新产品与技术的引入，令云数据中心的各类数据处理过程变得更为复杂。

这一趋势，伴随着数据规模的不断增大（端口带宽正逐步从 25G 向 100G 乃至更高演进），正对云数据中心带来巨大挑战，包括：

- 数据处理的增长速度始终高于算力增长速度，因此对处理器资源的占用率也越来越大。数据表明，目前数据中心中 30% 的计算是在作流量处理，甚至被形象地称为“数据中心税 (Datacenter Tax)”¹；
- 为保障核心业务的高效执行，数据中心不得不购置更多处理器资源，带来云服务总拥有成本（Total Cost of Ownership, TCO）的增加，进而抬升最终用户部署和使用云服务的成本；
- 更大的数据规模和处理复杂度，也对数据中心的各项基础设施性能，包括网络 I/O、转发、存储、安全以及管理等造成影响，阻碍云数据中心增加网络吞吐量、降低网络延迟等性能优化。

在这一背景下，不断革新的云数据中心亟待寻求一种全新的智能架构来化解上述挑战。对传统网卡产品进行“扩容赋能”，是业界解决数据处理负载过重的最初方案。这一被称为智能网卡

(SmartNIC) 的产品，通过在网卡中加入用于处理和转发不同端口上各种类型的数据等功能，从而实现更多的算力释放、更佳的网络性能表现，并降低处理延迟。但是这种智能网卡由于缺乏计算单元，往往只能做到数据平面的卸载，控制平面以及 Hypervisor 等基础设施管理负载仍然需要主机处理单元来完成，做不到基础设施处理的全卸载。

英特尔通过对数据中心基础设施能力进一步分析与整合，推出了英特尔® IPU (Intel® Infrastructure Processing Unit)，将数据平面，控制平面以及 Hypervisor 等基础设施管理负载和云服务基础设施能力都从主机服务器中卸载，一方面让宝贵的算力资源更多地用于用户关键业务，另一方面也通过将各种能力卸载到专用设备获得加速，有效提升各种数据处理负载的执行效率，来全面帮助云数据中心构建新的基础设施能力加速引擎。

解决方案：以英特尔® IPU，构建云数据中心加速引擎

IPU 是全新的产品形态。它在涵盖智能网卡功能的同时，更能对未来云数据中心带来更高级别的安全性和控制力。其优势包括：

- 卸载高密集的基础设施应用任务到 IPU 并进行加速，如加密和解密以及数据包处理；
- 在极端情况下，IPU 可以卸载整个虚拟机管理程序，释放处理器所有内核为应用和微服务提供支持，这对于裸金属服务产品来说至关重要；
- 减少了主机侧处理器中虚拟机管理程序和基础设施堆栈的开销，并可以从主机侧处理器中卸载存储堆栈，使更多的处理器资源可用于应用和租户工作负载；
- 使租户应用和基础设施管理解耦，实现更高安全性和可控性；
- 为裸金属和云主机统一云管控创造了可能。

如图1所示，英特尔® IPU 产品在配置了FPGA 芯片或者ASIC 中的专有硬件加速器之外，也加入通用处理器来对各类基础设施控制面功能实现卸载。这种将专用可编程硬件与通用处理器相结合的方式，不仅能使云数据中心中各基础设施能力获得预期加速，满足不断增长的网络数据处理需求，也能实现对控制面功能的灵活管理与操控，实现更优的系统级安全、控制和

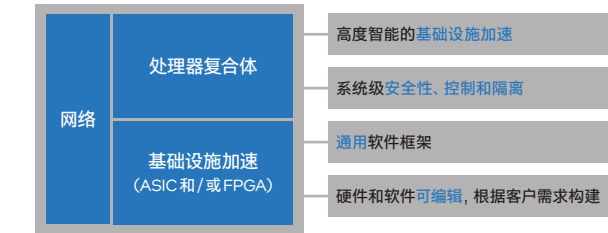


图1 英特尔® IPU 架构

以目前英特尔面向市场的 IPU 主力产品英特尔® FPGA IPU C5000X-PL (代号“Big Spring Canyon”) 为例，如图2所示，其具备2个25G 端口(使用 SFP28 光口)，可提供50G 的网络吞吐能力;核心处理能力则由所搭载的英特尔® Stratix® 10 DX FPGA 芯片与英特尔® 至强® D 处理器担纲。

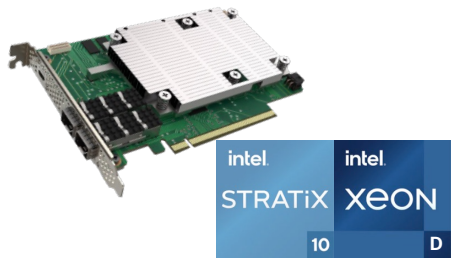


图2 英特尔® FPGA IPU C5000X-PL

其中，英特尔® Stratix® 10 DX FPGA 可编程逻辑芯片能充分发挥其硬件可编程特性，通过先进的架构设计、封装技术、相比前一代FPGA 拥有更多的收发器并支持硬核 PCIe Gen4 接口，从而实现更高的带宽，以定制化设计来实现高吞吐和低延时的性能表现，完成 I/O 虚拟化、OVS 转发等任务，实现基础设施管理、网络和存储功能的卸载。

英特尔® 至强® D 处理器则通过高度集成的设计，不仅能以卓越的单核性能来承载控制面的各项功能，支持 Hypervisor，还能以良好的 x86 兼容性以及与其它英特尔® 架构硬件形成的良好生态，帮助用户实现系统代码或应用能力的快速迁移，从而提升卸载效率。

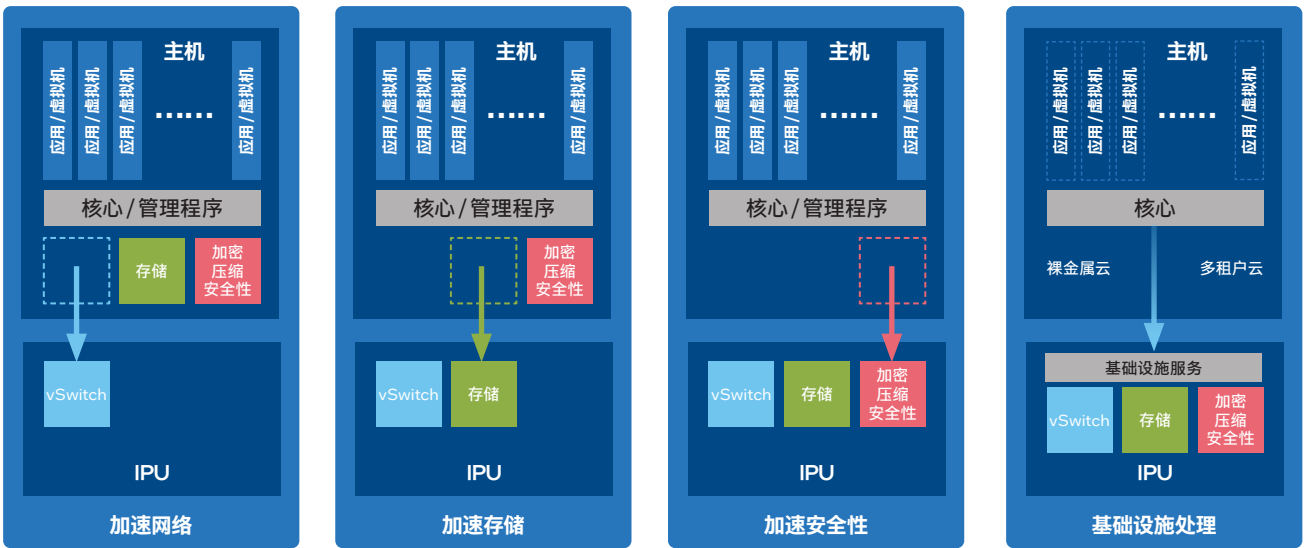


图3 英特尔® IPU 面向不同基础功能卸载的场景

■ 卸载基础设施能力与释放算力

借助上述两种芯片,如图 3 所示,云数据中心能够有效地将网络、存储、安全以及基础设施管理等多种能力从处理器卸载到 IPU,从而实现算力释放,并加速多种基础设施能力。这些能力包括:

- **网络加速:** 将承载网络 I/O、数据转发等功能的虚拟交换机软件,如 OVS 等从主机侧处理器卸载到 IPU,提升网络吞吐量,减少网络处理时延;
- **存储加速:** 将 virtio-blk、NVMe-oF 等存储接口、协议栈从主机侧处理器转移到 IPU,提高存储弹性和灵活性,并降低系统复杂性和开销;
- **安全加速:** 可以从主机侧处理器卸载加密/解密、压缩和其他应用功能;
- **基础设施处理:** 将云服务管理功能从主机侧处理器卸到 IPU,使虚拟机、容器或裸金属服务的分配和管理更为高效。

除了引入性能更强且具备灵活可编程特性的硬件来面向特定功能进行优化加速,释放更多宝贵算力之外,英特尔也赋予了 IPU 产品丰富的软件生态。目前,英特尔正以加速开发平台(Acceleration Development Platform, ADP)为抓手,优化生态系统,助力合作伙伴实现云数据中心性能加速解决方案的快速开发和部署。在规划中,ADP 平台将为用户提供通过英特尔® 开放式 FPGA 堆栈(英特尔® OFS)实现的板卡硬件设计、软件、驱动程序以及技术设计支持。用户可以利用这些软件工具与技术支持,加速 IPU 相关产品的开发,并迅速将产品推向市场。

■ 对接英特尔® 至强® 可扩展处理器,提升云数据中心算力

在英特尔为用户提供的高效能云数据中心生态方案中,不仅可以通过英特尔® IPU 开展能力卸载,对基础设施进行性能加速,也可以引入英特尔® 至强® 可扩展处理器来进一步加强算

力。通过对算力的“开源节流”,实现更快的数据处理速度、更大的带宽接入能力以及更低的网络时延。

第三代英特尔® 至强® 可扩展处理器可为服务器带来的性能增强包括:

- 更多的内核、更优的架构带来算力性能的大幅提升,可有效应对高密度计算所需;
- 支持更多内存,支持 PCIe Gen4,可实现更高的每核 I/O 带宽;
- 多项内置增强技术,如英特尔® 深度学习加速技术(英特尔® DL Boost)等可在人工智能等场景提供强大加速能力。

在算力架构中引入英特尔® 至强® 可扩展处理器的另一个优势是能够有效提升能力卸载的效率和平滑性。由于主机侧和 IPU 都采用基于英特尔® 架构的处理器(英特尔® 至强® 可扩展处理器和英特尔® 至强® D 处理器),这样将原先主机侧运行的应用迁移到 IPU 上就变的非常方便,甚至无需编译即可迁移。例如 DPDK、SPDK 等软件就可以从主机侧直接迁移到 IPU 上运行,从而大幅提升开发、部署和测试的效率。

这一算力架构正帮助云数据中心形成更高效的能力输出。以裸金属服务器为例,作为兼具物理机的高性能和高安全性,裸金属正越来越多地被部署在关键业务场景中。但传统上,由于裸金属的云管理能力由处理器承载,而裸金属又具有用户独占处理器资源的特性,因此裸金属无法像虚拟云主机那样便捷弹性地向用户提供算力。

借助英特尔® IPU 提供的云管理能力卸载,裸金属可以实现基础设施管理面与租户间的物理隔离,利用 VirtIO 设备热插拔特性支持弹性服务。而第三代英特尔® 至强® 可扩展处理器的加入,则进一步提升了裸金属中可供调度分配的算力,可以帮助云数据中心有能力为用户关键业务提供更具性能、更安全的服务,并提升算力资源的利用效率,从而提供兼具性能和性价比优势的云服务能力。

助力百度智能云打造更高效可靠的云服务能力

“云计算正与其它技术浪潮一起, 推动各行业的数字化、智能化转型。在这一进程中, 我们第一次自研推出 DPU 设备 —— 百度太行 DPU 1.0。它能够支持太行弹性裸金属服务, 实现网络卸载、PCIe 设备热插拔等功能, 让云计算的虚拟化开销很大程度沉淀在网卡上, 使云产品的性能大大提升。”

侯震宇
百度集团副总裁

作为领先的云服务提供商, 百度智能云在云数据中心建设方向上与英特尔有着同样的理念, 即应该将基础设施能力的相关开销全部从主机侧处理器卸载到专门的基础设施处理单元上并实施性能加速, 以便释放算力, 优化整体性能, 从而打造高效、安全和可编程的未来数据中心。

来自百度智能云的实践表明, 持续向各个领域延伸, 并承载越来越多核心业务能力的云服务, 正面临着需要更强算力资源、更大带宽接入和更低网络时延的挑战。以百度智能云独具优势的 AI 云服务为例, 目前 AI 算力需求比之前已提升数个数量级, 元宇宙的算力需求还要再提升 1000 倍²。因此要贯彻上述理念, 需要更为强劲的技术底座予以支持。

为此, 百度智能云致力于发展软硬件结合的云计算基础设施, 通过自研的百度太行 DPU 系列产品, 并融合其在云服务领域的丰富经验以及大量真实需求, 来应对云服务实践中遇到的算力、网络、存储、安全等一系列挑战。

如下页图 5 和 表 1 所示, 百度太行 DPU 1.0 产品配备了 2 个 25G 带宽的光口, 并引入英特尔® FPGA 与英特尔® 至强® D 处理器作为算力核心, 来实现各项虚拟化功能的卸载, 在为户提供弹性网卡、弹性存储能力之余, 还支持 1024 个设备的热插拔。同时, 通过将云管控平面卸载到百度太行 DPU 1.0 中, 也可以实现虚拟云主机和裸金属共池, 即裸金属可与虚拟云主机被统一灵活调度。来自百度智能云的测试验证表明, 百度太行 DPU 1.0 产品能够为用户提供 1000 万 PPS 的转发率以及 20 万 IOPS 的存储性能³。

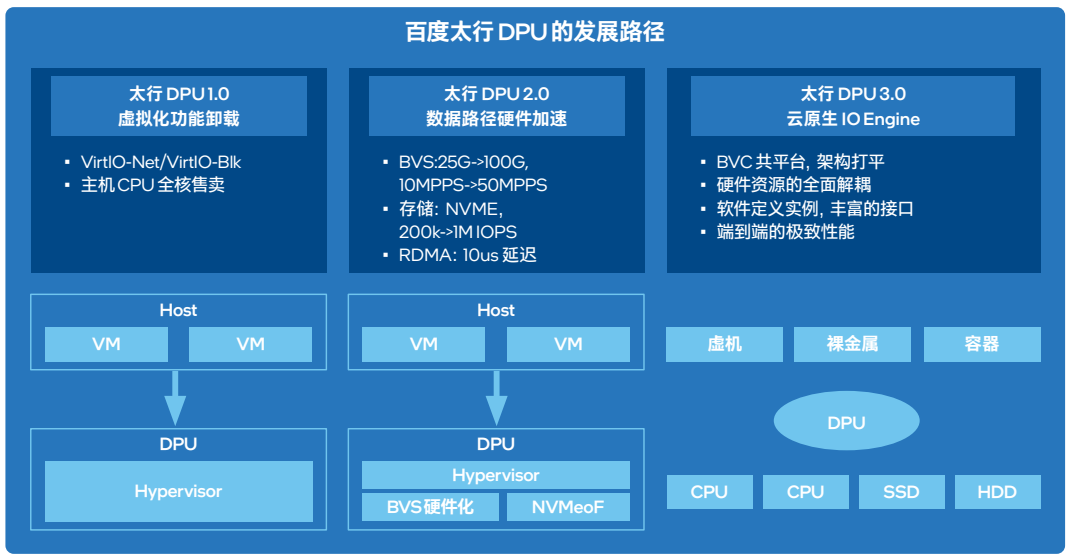


图 4 百度太行 DPU 发展路径



图5 百度太行 DPU 1.0 产品图

百度太行 DPU 产品功能亮点与配置数据
功能亮点—使能弹性裸金属、虚拟机全场景部署
<ul style="list-style-type: none">▪ 支持网络卸载及加速, 实现裸金属和虚拟机物理机型统一;▪ 支持存储卸载及加速, 让裸金属可以和虚拟机一样实现云盘快速部署和云盘数据盘, 以及执行磁盘快照、数据疏散等能力;▪ 支持弹性热插拔, 为裸金属和虚拟机提供弹性网卡、弹性扩容等关键能力;▪ 支持热升级、热恢复、热迁移(虚拟机)。
百度太行 DPU 1.0—虚拟化功能卸载
<ul style="list-style-type: none">▪ 弹性网卡, 弹性存储, 1024 个设备热插拔▪ 云管控平面卸载到 DPU, 虚拟机和裸金属共池▪ 2*25G, 10MPPS, 200K IOPS

表1 百度太行 DPU 产品功能亮点与配置数据⁴

现在, 百度太行 DPU 1.0 产品已在太行弹性裸金属等多种云服
务场景中开展应用并获得了用户好评。实践证明, 百度自研的
百度太行 DPU 1.0 可以有效解决云数据中心管理及虚拟化的

局限性, 帮助百度智能云保持优势。在百度智能云的计划中,
这一 DPU 产品也将作为其智能云服务的核心组件, 助力其构
建基础设施即服务 (Infrastructure as a Service, IaaS) 资源
的统一弹性底座, 进而为用户依托百度智能云开展各类业务创
新、实施产业智能化转型提供高效且可靠的性能保障。

总结与展望

随着英特尔® IPU 以及基于其参考设计研发的相关产品在云
数据中心的基础架构革新中展现出显著性能优势和市场价
值, 并逐步形成可持续发展的生态链, 英特尔与百度智能云等
深度合作伙伴也正围绕云服务的未来发展方向, 从软硬件层
面开拓更多优化方案。例如通过新的英特尔® IPU 平台提供更
大网络吞吐量, 实现对基础设施管理的全面解耦, 以及以极低
时延实现端到端的加速等。

面向未来, 英特尔还计划在 IPU 产品体系上持续发力。如图 6
所示⁵, 英特尔在 2022 年将推出两款第二代 200G 的可编程
IPU, 代号分别为“Oak Springs Canyon”和“Mount Evans”。
其中, Oak Springs Canyon 搭载了英特尔® 至强® D 处理器
与英特尔® Agilex™ FPGA 芯片; 而 Mount Evans 则是英特尔
首款基于 ASIC 芯片的 IPU 产品。这两款产品都将具有卸载

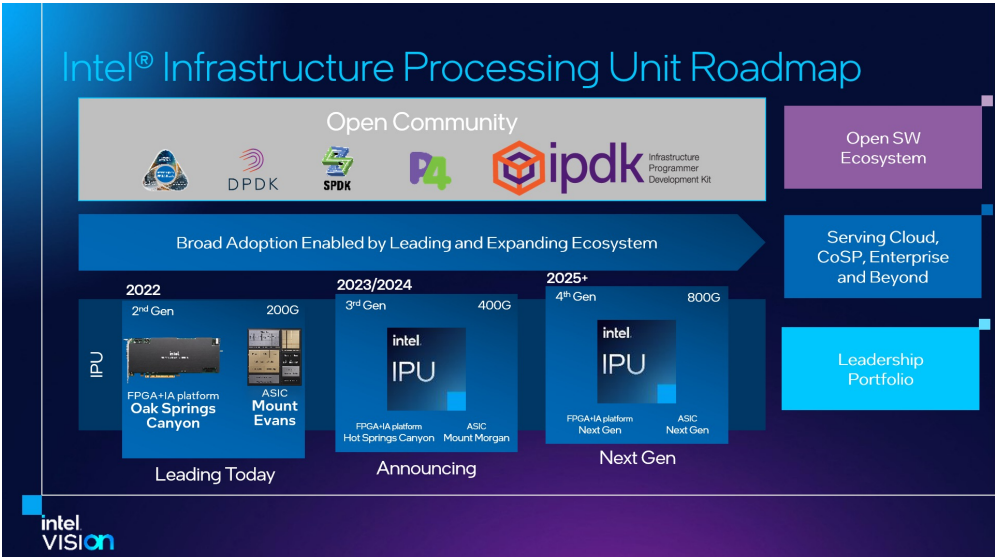


图6 英特尔® IPU 产品路线图

200G 工作负载所需的基础设施加速能力,且都可与英特尔®至强®可扩展处理器协同工作。未来两到三年内,英特尔将推出支持 400G 的 IPU 产品,在 2025 年或晚些时候,推出新一代基于 FPGA 芯片和 ASIC 芯片的 800G IPU 产品。

与此同时,英特尔也正进一步背靠丰富的开源软件,如基础架构编程人员开发套件 (Intel®oneAPI™) 和

Development Kit, IPDK)、SPDK、DPDK 等,致力于通过开放的生态系统、多方位的行业协作以及积极的社区互动,让百度智能云等云服务提供商、云服务最终用户以及其它合作伙伴能加速 IPU 相关解决方案的构建,从云数据中心基础架构的持续革新中撷取更大价值。



¹该观点援引自公开媒体发布的《从 DPU 的崛起谈谈计算体系变革》一文: <https://aijishu.com/a/1060000000228825>。

²数据援引自公开媒体报道: <https://finance.sina.com.cn/chanjing/cywx/2022-03-10/doc-imcwiwss5271286.shtml>

³百度太行 DPU 产品测试数据来源于百度内部

⁴百度太行 DPU 产品数据来源于百度内部

⁵资料援引自文章《英特尔在 2022 英特尔 on 产业创新峰会上推出多代基础设施处理器 (IPU) 路线图》<https://www.intel.cn/content/www/cn/zh/now/data-centric/rolls-out-multi-generation-ipu-roadmap-at-vision.html>

实际性能受使用情况、配置和其他因素的差异影响。更多信息请见 www.intel.com/PerformanceIndex

本文并未 (明示或默示、或通过禁止反言或以其他方式) 授予任何知识产权许可。英特尔未做出任何明示和默示的保证,包括但不限于,关于适销性、适合特定目的及不侵权的默示保证,以及在履约过程、交易过程或贸易惯例中引起的任何保证。

英特尔运营所需的任何商品和服务预测仅供讨论。就与本文中公布的预测,英特尔不负有任何购买责任。本文中提供的所有信息可在不通知的情况下随时发生变更。关于英特尔最新的产品规格和路线图,请联系您的英特尔代表。

英特尔技术特性和优势取决于系统配置,并可能需要支持的硬件、软件或服务得以激活。产品性能会基于系统配置有所变化。没有任何产品或组件是绝对安全的。更多信息请从原始设备制造商或零售商处获得,或请见 [intel.com](https://www.intel.com)

在特定系统的特殊测试中测试组件性能。硬件、软件或配置差异将影响实际性能。当您考虑采购时,请查阅其他信息来源评估性能。关于性能和基准测试程序结果的更多信息,请访问: www.intel.com/benchmarks

英特尔并不控制或审计第三方数据。请您自行审核该内容、咨询其他来源,并确认提及数据是否准确。

©英特尔公司版权所有。英特尔、英特尔标识以及其他英特尔商标是英特尔公司或其子公司在美国和/或其他国家的商标。