

业务简介

第四代英特尔® 至强® 可扩展处理器
云计算



为您的云增长提供无障碍环境



借助英特尔® 技术，您可以启动成熟且灵活的混合云、多云或智能边缘策略。

随着云迁移不断在企业范围内展开，云服务环境的情况变得越来越复杂。由行业最佳实践驱动的同步化业务与技术策略，可以提升由云优先运营带来的灵活性、成本效益和其他优势。企业和机构既需要提升效率，提高服务水平的可预测性，同时又要加速创新以及更好地支持业务决策。

在企业和机构寻求扩大规模、降低成本和提供新服务的过程中，通过技术来实现商业价值的重要性日益凸显。面对新的应用场景，他们无需定制系统（这可能会增加复杂性），而是可以通过使用易于扩展的平台来满足当下和未来各种部署的性能需求。

89% 拥有多云策略¹
的企业和机构

16.9% 全球全云增长²
的 CAGR (到 2025 年)

1.3 万亿美元 全球全云支出²
(到 2025 年)

运用英特尔® 技术加速上云

第四代英特尔® 至强® 可扩展处理器内置众多加速器，可为 AI、数据分析、网络、存储和科学计算等快速增长的云工作负载提供性能和能效优势。在全新英特尔® 高级矩阵扩展 (Intel® Advanced Matrix Extensions, 英特尔® AMX) 这一加速器的支持下，第四代英特尔® 至强® 可扩展处理器具有更为出色的 AI 训练和推理性能。而其他无缝内置于处理器中的加速器则可分别加速数据传输和压缩，从而实现更高的网络性能；可提升查询吞吐量，从而实现响应更加迅捷的数据分析；同时还可将调度和队列管理任务卸载到云平台的多个内核上，实现工作负载动态平衡。为在超大规模环境下实现新的内置加速器功能，英特尔为生态系统提供了各种常见的云 API、库和操作系统级软件支持。

通过内置加速器和软件优化，上一代英特尔® 至强® 可扩展处理器已被证明可以在真实场景下的目标云工作负载上实现出色的性能功耗比³。这不但可以提高 CPU 利用率，降低云功耗，并提高服务的投资回报率 (ROI)，同时还能帮助企业实现可持续发展目标。

性能证明

高达 **2.5 倍** 的吞吐量
提升 (RPS)

高达 **74%** 的 P99 时延
降幅

高达 **12%** 的 CPU 利用率
降幅

数据基于内置两个英特尔® QAT 的英特尔® 至强® 铂金 8480+ 处理器与未使用加速技术的解决方案的比较⁴





关键云工作负载

机密计算

企业和机构十分注重严格的安全性、身份与合规管理，这是他们零信任安全策略的一部分。机密计算通过基于硬件的内存保护功能提升敏感数据的独立性。到 2026 年，机密计算市场的复合年增长率 (CAGR) 预计将达 90%-95%，达到 540 亿美元⁵。

需求:

出于隐私、监管或知识产权方面的考虑，与第三方共享数据往往非常困难，甚至不可能实现。如果能克服这些障碍，便能获得更多潜在好处，例如使用大规模分布式数据集支持研究或训练 AI 模型，充分利用云在成本、可扩展性和敏捷性上的优势。

答案:

采用第四代英特尔® 至强® 可扩展处理器实现机密计算，通过在基于硬件级保护的内存飞地内只允许受信任的代码访问数据来提高数据的独立性。这些隐私保护功能使超大规模平台和其他分布式网络中围绕受监管工作负载的创新成为可能。



- **英特尔® 软件防护扩展 (Intel® Software Guard Extensions, 英特尔® SGX)** 是目前市场上经过深入研究、多次更新和广泛部署的数据中心级机密计算技术，拥有极小的信任边界。机密计算通过基于硬件的内存保护功能提升敏感数据的独立性。
- **英特尔® Trust Domain Extensions (英特尔® TDX)** 帮助保护虚拟机 (VM) 内的数据访问，只需轻点几下，便可以仅允许 VM 中的软件访问 VM 中的数据。

微服务

微服务的采用仍然是开发型企业内部转型的主要推动因素，特别是对超大规模平台和智能边缘环境来说。英特尔最近开展的一项研究显示，超过 83% 的全新云原生应用以及软件即服务 (SaaS) 解决方案都在使用微服务⁶。

需求：

企业需要在跨云环境中无缝运营，确保他们的应用能在大规模环境下运行得更好、更快。云硬件和软件必须针对吞吐量进行优化以支持应对关键的微服务挑战，例如服务质量、基础设施开销和可观测性。

性能证明

高达 的吞吐量
1.88 倍 提升

数据基于第四代英特尔® 至强® 可扩展处理器与上一代产品在 DeathStarBench [社交网络微服务工作负载 (读取主页时间线)，100ms SLA] 上的表现的比较⁷

答案：

第四代英特尔® 至强® 可扩展处理器内置的加速器可以加速云平台间和云平台内的数据传输，从而保障服务质量、基础设施开销和数据/流程可观测性方面运营目标的实现。它们有助于确保快速启动和关闭云业务服务，支持硬件辅助队列，以帮助数百万计的操作请求及高速分布式网络通信实现负载平衡。

- **英特尔® 数据流加速器 (Intel® Data Streaming Accelerator, 英特尔® DSA)** 可通过优化流数据的传输和转换操作，大幅提升存储、网络和数据密集型工作负载的性能。常见的数据传输任务会导致数据中心级部署产生开销。为卸载这些任务而设计的英特尔® DSA 有助于加速数据在 CPU、内存和缓存以及各种附加内存、存储和网络设备中的传输。
- **英特尔® 动态负载均衡器 (Intel® Dynamic Load Balancer, 英特尔® DLB)** 可在多核英特尔® 至强® 可扩展处理器上提升与网络数据处理相关的系统性能。英特尔® DLB 能够在多个 CPU 内核/线程之间有效地分配网络处理负载，并能随着系统负载的变化，将网络数据动态地分配到多个 CPU 内核上进行处理。此外，英特尔® DLB 还可以复位同时在 CPU 内核上处理的多个网络数据包的顺序。

性能证明

高达 的时延
96% 降幅

数据基于在相同吞吐量 (RPS) 下，第四代英特尔® 至强® 可扩展处理器启用英特尔® DLB 与在 6 核 (12 线程) 处理器上运行软件来处理 Istio 入口网关连接请求的比较⁸

性能证明

高达 **2.33 倍** 提升 与 高达 **2.63 倍** 提升
[MTU 1500, 1500B] [MTU 9000, 2400B]

数据基于第四代英特尔® 至强® 可扩展处理器启用英特尔® DSA 进行内存拷贝与使用软件进行内存拷贝的比较⁹

AI 与高级数据分析

利用 AI 和数据分析的融合能够帮助企业了解自身的数据，获得运营洞察并在数据驱动下做出战略决策。84% 的企业高管认为他们需要运用 AI 来实现增长目标，但 76% 的企业高管承认，他们在如何扩展企业和服务中的 AI 规模时遇到了挑战¹⁰。

需求：

随着各类不同规模的企业越来越多地运用 AI 和数据分析，他们必须优化性能和其他具有积极投资回报的运营成果。

答案：

借助第四代英特尔® 至强® 可扩展处理器，企业能够从数据中获取更多的洞察和价值，从而利用 AI 做出数据驱动型决策。金融服务机构可以做出更好的投资和风险管理决策。医疗保健服务企业和机构可以通过更高效的计费 and 预先审批流程以及更准确的术后并发症预测，来改进患者护理并降低成本。线上以及零售企业将获得对特定客户需求的独特理解和联系，并通过更好的需求预测、主动和个性化的营销以及实时物流和库存管理，提高营收。

- **英特尔® 高级矩阵扩展 (英特尔® AMX)** 强化了第四代英特尔® 至强® 可扩展处理器的 AI 加速能力，无需额外硬件即可加速训练和推理。该内置加速器可为自然语言处理、推荐系统和图像识别等 AI 应用提供更强的支持。
- **英特尔® 高级矢量扩展 512 (Intel® Advanced Vector Extensions 512, 英特尔® AVX-512)** 可为密集型工作负载提升性能。它是最新的 x86 矢量指令集，包含多达两个融合乘加 (FMA) 单元和其他优化，可为要求严苛的计算任务提升性能。

性能证明

高达 **5.7 倍至 10 倍** 的 PyTorch 实时推理性能提升¹¹ 高达 **3.5 倍至 10 倍** 的 PyTorch 训练性能提升¹²

数据基于启用内置英特尔® AMX (BF16) 的第四代英特尔® 至强® 可扩展处理器与上一代产品 (FP32) 的比较

提高工作效率

提升跨多个计算节点的超大规模云服务和应用的性能。

[Google](#)、[Meta](#)、[Amazon](#) 和 [Microsoft](#) 等重要客户都认识到了大幅提升跨数据中心的时间同步准确性的价值，其中很多客户已经部署了依赖精确网络时间同步（如 IEEE 1588-2019 标准等）的分布式应用。第四代英特尔® 至强® 可扩展处理器采用精确时间管理 (PTM) 机制，它在服务器内将精确的网络时间从 PCIe 网络设备一直扩展到 CPU 上运行的应用软件，从而完成服务器之间可实现的数量级性能提升。

有关 PTM 的更多信息，请参阅[“采用时间敏感网络 \(TSN\) 实现自动化系统。”](#)

利用既有基础设施轻松集成

与英特尔合作，企业可以利用他们已经熟悉和正在使用的大规模合作伙伴生态系统缩短部署时间。全球各地的硬件和软件供应商以及解决方案集成商都在使用英特尔® 至强® 可扩展处理器构建其产品，并通过数以千计来自真实场景的实现案例提供更多选择和更好的互操作性。

数十年积累下来的生态系统支持有助于英特尔将其可信计算基础扩展到超大规模数据中心和新的边缘环境之中，并且使英特尔有能力去推进更多构建、扩展和转型工作，帮助企业实现更优的运营敏捷性。这种生态系统上的开放性使得用户可在硬件、软件、云和服务提供商中自由选择。通过[英特尔® 合作伙伴联盟](#)，您可访问 AI、云、科学计算和其他解决方案领域的更多专属资源，帮助您规划、构建并为客户提供更多价值。

英特尔® 架构具有很强的灵活性，可在其上选择不同的云服务，并对其进行进一步的量身定制，以支持特定工作负载的需求，因此其应用规模可扩展到全球范围，并覆盖各个主流云服务提供商。

支持性统计数字

利用英特尔 **超 50,000 个** 独特实例类型、规格和区域，获得更多选择。

领导层在数字化转型之旅中的首要业务目标

2022 至 2024 年间，企业和机构领导层（科技和商业公司）在数字化转型方面的投资预计将达到 6.3 万亿美元，到 2024 年，这一数字将占有所有 IT 支出的 55%¹³。本业务简介是一系列业务简介的一部分，旨在阐明领导者在未来的变革中实现业务成功所关注的首要业务目标，以及英特尔® 硬件、软件和服务（包括第四代英特尔® 至强® 可扩展处理器）将如何帮助实现这些目标：



- **云（此简介）：**启动跨混合云、多云与智能边缘的策略
- **安全性：**实现更严格的安全性，推动零信任安全策略
- **AI：**采用数据分析和 AI 来驱动关键成果的产出
- **重新定义员工体验：**支持无边界互动式员工体验
- **ESG：**促进环境中的公平结果与责任 | 社会 | 治理 (ESG)

了解更多信息

www.intel.cn/xeon/scalable
www.intel.cn/cloud



¹ Flexera, “2022 State of the Cloud Report” (2022 年云状态报告), <https://resources.flexera.com/web/pdf/Flexera-State-of-the-Cloud-Report-2022.pdf>.

² IDC 市场预测, 2021 年 9 月 (IDC #US47397521), “Worldwide Whole Cloud Forecast, 2021–2025: The Path Ahead for Cloud in a Digital-First World” (全球全云预测, 2021–2025 年: 数字优先世界中的云未来之路), <https://www.idc.com/getdoc.jsp?containerId=US47397521>.

³ 配置详情请见以下网址的 [I26-I30]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/3rd-generation-intel-xeon-scalable-processors/>.

⁴ 详情请见以下网址的 [W5]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/processors/> (第四代英特尔® 至强® 可扩展处理器)。结果可能不同。

⁵ Everest Group, 2021 年 10 月, “Confidential Computing – The Next Frontier in Data Security” (机密计算——数字安全的下一个前沿), https://confidentialcomputing.io/wp-content/uploads/sites/85/2021/10/Everest_Group_-_Confidential_Computing_-_The_Next_Frontier_in_Data_Security_-_2021-10-19.pdf.

⁶ 英特尔微服务洞察研究, 2021 年 6 月。

⁷ 请参阅 [W1], 网址为 intel.com/processorclaims: 第四代英特尔® 至强® 可扩展处理器。结果可能不同。

⁸ 详情请见以下网址的 [W6]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/processors/> (第四代英特尔® 至强® 可扩展处理器)。结果可能不同。

⁹ 详情请见以下网址的 [W7]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/processors/> (第四代英特尔® 至强® 可扩展处理器)。结果可能不同。

¹⁰ Accenture, 2019 年 11 月 19 日, “AI: Built to Scale” (AI: 为扩展而构建), <https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments>.

¹¹ 详情请见以下网址的 [A17]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/processors/> (第四代英特尔® 至强® 可扩展处理器)。结果可能不同。

¹² 详情请见以下网址的 [A16]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/processors/> (第四代英特尔® 至强® 可扩展处理器)。结果可能不同。

¹³ IDC, 2021 年 10 月, “IDC FutureScape: Worldwide Digital Transformation 2022 Predictions” (IDC FutureScape: 2022 年全球数字化转型预测), <https://www.idc.com/getdoc.jsp?containerId=US47115521>.

加速器是否可用视 SKU 而定。更多产品详情, 请见英特尔® 产品规格页面。

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 <https://intel.cn/PerformanceIndex>。

性能测试结果基于配置信息中显示的日期进行的测试, 且可能并未反映所有公开可用的安全更新。详情请参阅配置信息披露。没有任何产品或组件是绝对安全的。

英特尔并不控制或审计第三方数据。请您审查该内容, 咨询其他来源, 并确认提及数据是否准确。

具体成本和结果可能不同。

英特尔技术可能需要启用硬件、软件或激活服务。

您不得将此文件用于或协助用于任何关于英特尔产品的侵权或其他法律分析的文件。对于后续起草的包含本文所披露标的物的任何专利权利要求, 您同意授予英特尔非排他的、免许可费的许可。

描述的产品可能包含可能导致产品与公布的技术规格有所偏差的、被称为非重要错误的设计瑕疵或错误。一经要求, 我们将提供当前描述的非重要错误。

© 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的商标。其他的名称和品牌可能是其他所有者的资产。

1122/MH/MESH