

第四代英特尔® 至强® 可扩展处理器



众多内置加速器，让性能更进一步

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现在和未来各种部署的性能需求。

第四代英特尔® 至强® 可扩展处理器旨在为人工智能 (AI)、数据分析、存储和科学计算方面快速增长的工作负载提供性能加速。该处理器具备多种内置加速器，帮助人们将零信任安全策略付诸实践，同时利用先进的安全技术，即使面对敏感或受监管的数据，也能解锁新的商业合作机会和洞察。使用这款处理器，您可以跨多个云和边缘环境进行扩展，满足自身的部署需求。英特尔® 至强® 可扩展处理器具有很强的灵活性，可在其上选择不同的云服务，帮助企业顺利实现应用移植。

英特尔几大加速引擎，重新定义性能

重新定义您对处理器的期望。与增加 CPU 内核数相比，内置加速器是一种提升性能更有效的方法。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工作负载上实现出色的性能功耗比¹。这不但可以提高 CPU 利用率，降低功耗，并提高投资回报率 (ROI)，同时还能帮助企业实现可持续发展目标。

- **英特尔® 高级矩阵扩展 (Intel® Advanced Matrix Extensions, 英特尔® AMX)** 可加速自然语言处理 (NLP)、推荐系统和图像识别等深度学习 (DL) 推理和训练工作负载。
- **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。
- **英特尔® 存内分析加速器 (Intel® In-Memory Analytics Accelerator, 英特尔® IAA)** 可提高数据分析性能，同时从 CPU 内核上卸载任务，为数据库查询及其他工作负载加速。
- **英特尔® 动态负载均衡器 (Intel® Dynamic Load Balancer, 英特尔® DLB)** 可随系统负载的变化将网络数据动态地分配到多个 CPU 内核上，基于硬件高效实现负载平衡。
- **面向 vRAN 的英特尔® 高级矢量扩展 (Intel® Advanced Vector Extensions, 英特尔® AVX)** 在相同功耗范围内可将虚拟无线接入网络 (vRAN) 的密度较上一代产品提高多达 2 倍²。
- **英特尔® 高级矢量扩展 512 (Intel® Advanced Vector Extensions 512, 英特尔® AVX-512)** 支持多达两个融合乘加 (FMA) 单元，并包含多项优化，可为要求严苛的计算任务提高性能。
- **英特尔® 数据保护与压缩加速技术 (英特尔® QAT)** 可加速加解密和数据压缩，它通过从处理器内核卸载这些任务，帮助降低系统资源消耗。
- **英特尔® 密码操作硬件加速 (Intel® Crypto Acceleration)** 降低了实施普遍数据加密的影响，并提高了安全套接字层 (SSL) Web 服务器、5G 基础设施和 VPN/防火墙等加密敏感型工作负载的性能。

AI

凭借更优的矢量指令和矩阵乘法运算，第四代英特尔® 至强® 可扩展处理器展现出更为出色的 AI 推理和训练性能。英特尔® AMX 可以显著提高推荐系统、NLP、图像识别、媒体处理和分发以及媒体分析等深度学习工作负载的性能。

科学计算

第四代英特尔® 至强® 可扩展处理器可以提高科学计算工作负载中常见的多线程代码性能。这些工作负载包括制造业仿真、分子动力学、地球系统建模以及 AI 推理和训练。该处理器内置的加速器可提供较高的精度水平，同时还能加速多种 AI 数据类型的处理。该处理器还支持 DDR5 内存、PCIe Gen5、英特尔® 超级通道互联 (Intel® Ultra Path Interconnect, 英特尔® UPI) 2.0 和 Compute Express Link (CXL)，显著提升了整体数据的吞吐量。

数据分析

第四代英特尔® 至强® 可扩展处理器内置的加速器可增强内存数据库、大数据、数据仓库、商业智能 (BI)、企业资源规划 (ERP) 和运营分析的性能。其中，英特尔® DSA 可改善数据处理密集型应用中常见的流数据传输和转换操作，英特尔® IAA 可通过卸载 CPU 内核的任务来提升数据库操作的吞吐量。

网络和存储

英特尔® DLB 可平衡内核间的操作并实现网络数据包优先排序。英特尔® DSA 可卸载数据复制及常见的数据转换操作，从而释放内核周期。这些内置加速器能够实现高效的网络数据分配和企业存储数据传输，并能够改善云计算的内存管理操作，因而使云计算性能更为出色。

加密

英特尔® QAT 现已内置于第四代英特尔® 至强® 可扩展处理器，可为密码操作和压缩操作加速。英特尔® QAT 可以显著提高 CPU 效率和应用程序吞吐量，同时减少数据占用空间和功耗，赋能企业和机构在不影响性能的情况下强化加密操作。

安全性

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。

高达 **1.53 倍**

平均性能增益
(与上一代产品的比较)³

高达 **10 倍**

PyTorch 实时推理和训练性能提升
这是在启用内置的英特尔® AMX (BF16) 时
与上一代产品 (FP32) 的比较结果⁴

高达 **3 倍**

RocksDB 性能提升
这是在启用内置的英特尔® IAA 时
与上一代产品的比较结果⁵

高达 **1.6 倍**

的 IOPS 提升和高达 37% 的时延降低
这是面对大型数据包顺序读任务，
启用英特尔® DSA 时与上一代产品的比较结果⁶

高达 **2 倍**

容量增益
这是在相同功耗范围内与上一代产品
在 vRAN 工作负载方面的比较结果²

高达 **95%**

内核减少幅度以及 2 倍一级压缩吞吐量提升
这是启用内置的英特尔® QAT 时
与上一代产品的比较结果⁷

技术概述

第四代英特尔® 至强® 可扩展处理器采用全新架构，单核性能比上一代产品更高，每路配备多达60个内核。每个系统支持单路、双路、四路或八路配置。为了与内核数增加这种情况相匹配，该平台在内存和输入/输出 (I/O) 子系统方面也做了相应改进。DDR5 内存提供的带宽和速度与 DDR4 相比提高多达 1.5 倍，速率达到 4,800 MT/s⁸。此外，该平台还具有每路 80 条 PCIe Gen5 通道的特点，与之前的平台相比，I/O 得到显著提升⁹。本代处理器还可提供 CXL 1.1 连接，支持高网络带宽并使附加加速器能够高效运行。第四代英特尔® 至强® 可扩展处理器支持的技术让您可以根据工作负载要求的变化灵活扩展和调整。此外，本代处理器还可助您实现以下优势：

- 进一步提升网络、存储和计算性能，并通过将繁重的任务卸载到英特尔® 基础设施处理单元 (Intel® Infrastructure Processing Unit, 英特尔® IPU) 来提高 CPU 利用率
- 通过英特尔® UPI 2.0 提高多路带宽 (高达 16 GT/s)
- 使用英特尔® Speed Select 技术 (英特尔® SST) 调整 CPU 配置，满足特定工作负载的需求
- 增加三级缓存 (LLC) 共享容量 (所有内核共享多达 100 MB LLC)
- 通过硬件增强型安全功能加强您对安全态势的掌控
- 使用英特尔® Virtual RAID on CPU (英特尔® VROC) ，从而无需再用单独的 RAID 卡

第四代英特尔® 至强® 可扩展处理器的新特性或新功能

PCI Express Gen5 (PCIe 5.0)

带来全新的 I/O 速度，可在 CPU 和互联设备之间实现更高的吞吐量。第四代英特尔® 至强® 可扩展处理器具有多达 80 条 PCIe 5.0 通道，非常适合高速网络、高带宽加速器和高性能存储设备。PCIe 5.0 的 I/O 带宽是 PCIe 4.0 的两倍⁹，仍具备向后兼容性并提供用于 CXL 连接的基础插槽。

DDR5

以更高内存带宽克服数据瓶颈，提高计算性能。与 DDR4 相比，DDR5 的带宽提高多达 1.5 倍¹⁰，因此有机会提升性能、容量和能效并降低成本。借助 DDR5，第四代英特尔® 至强® 可扩展处理器提供的速率可高达 4,800 MT/s (1 DPC) 或 4,400 MT/s (2 DPC)。

CXL

借助面向下一代工作负载的 CXL 1.1，降低数据中心的计算时延并帮助减少总体拥有成本 (TCO)。CXL 是另一种跨标准 PCIe 物理层运行的协议，可以在同一链路上同时支持标准 PCIe 设备和 CXL 设备。CXL 可带来的一大关键能力是：在 CPU 和加速器之间创建统一且一致的内存空间，它将革新未来数年数据中心服务器架构的构建方式。

以丰富的选择和出色的灵活性实现扩展——英特尔® 至强® 可扩展处理器

从硬件到系统再到软件，英特尔提供值得信赖的技术基石，帮助企业和机构实现一系列不断扩张的业务目标，同时更好地保护数据安全。无论是提供更大的计算密度来降低功耗，还是加速 AI 工作流程，抑或是支持转向云原生架构，英特尔® 至强® 可扩展处理器都可帮助企业 and 机构解决至关重要的业务挑战，提供更丰富的云选择和更出色的应用可移植性。

第四代英特尔® 至强® 可扩展处理器概述

英特尔® 至强® 铂金 8400 处理器是打造安全且敏捷的混合云数据中心的基石，专为高级数据分析、AI、高密度基础设施和多云工作负载而设计。这些处理器具备更高性能、强大的平台功能和出色的工作负载加速能力。它们具有更出色的基于硬件的安全性和强大的多路处理性能——特定型号的英特尔® 至强® 铂金 8400 处理器支持多达 8 路配置。借助值得信赖且经过硬件增强的数据服务交付以及全新的 I/O 和连接技术，这些处理器在 I/O、内存、存储和网络技术方面均实现提升，因而能够更好地在数据驱动程度日益加深的世界中挖掘可执行洞察。相关提升包括：

- 每个英特尔® 至强® 可扩展处理器具备多达 60 个内核
- 每个处理器有 8 条内存通道，速率高达 4,800 MT/s (1DPC)
- 英特尔® AMX 支持下的 AI 加速带来深度学习推理和训练性能的巨大飞跃

英特尔® 至强® 金牌 6400 和英特尔® 至强® 金牌 5400 处理器最多可扩展到四路¹⁾，并已针对要求严苛的主流数据中心、多云计算以及网络和存储工作负载进行了优化。这些处理器支持更快的内存速度和更大的内存容量，带来了更高的性能、更强大的内存功能、硬件增强型安全性和工作负载加速。

英特尔® 至强® 银牌 4400 处理器可提供基本性能、更快的内存速度和更出色的能效，具备入门级数据中心计算、网络和存储所需的硬件增强型性能。

了解更多信息

有关上述处理器如何推动企业业务发展的更多信息，请见 intel.cn/xeonscalable 和 intel.cn/content/www/cn/zh/products/docs/processors/xeon-accelerated/4th-gen-xeon-scalable-processors.html。



多达 8 路的可扩展性

4 个英特尔® UPI 端口，速率为 16 GT/s

80 条 PCIe 5.0 通道 + CXL

支持 DDR5，速率高达 4,800 MT/s (每通道 1 个 DIMM) 或 4,400 MT/s (每通道 2 个 DIMM)

支持英特尔® 傲腾™ 持久内存 300 系列

英特尔® AVX-512 (两个 512 位 FMA)

英特尔® 超线程技术和英特尔® 睿频加速技术

英特尔® AMX

英特尔® SST

先进的可靠性、可用性和可维护性 (RAS)

英特尔® SGX 最大飞地容量高达 128 GB (在特定型号的 SKU 上最大飞地容量高达 512 GB)

可通过英特尔® QAT、英特尔® DLB、英特尔® DSA 和英特尔® IAA 加速工作负载

多达 4 路的可扩展性

3 个英特尔® UPI 端口，速率为 16 GT/s

80 条 PCIe 5.0 通道 + CXL

支持 DDR5，速率高达 4,800 MT/s (每通道 1 个 DIMM) 或 4,400 MT/s (每通道 2 个 DIMM)

支持英特尔® 傲腾™ 持久内存 300 系列

英特尔® AVX-512 (两个 512 位 FMA)

英特尔® 超线程技术和英特尔® 睿频加速技术

英特尔® 深度学习加速技术和英特尔® AMX

英特尔® SST

先进的 RAS

英特尔® SGX 最大飞地容量高达 128 GB

可通过英特尔® QAT、英特尔® DLB、英特尔® DSA 和英特尔® IAA 加速工作负载

多达 2 路的可扩展性

2 个英特尔® UPI 端口，速率为 16 GT/s

80 条 PCIe 5.0 通道 + CXL

支持 DDR5，速率高达 4,800 MT/s (每通道 1 个 DIMM) 或 4,400 MT/s (每通道 2 个 DIMM)

英特尔® AVX-512 (两个 512 位 FMA)

英特尔® 超线程技术和英特尔® 睿频加速技术

英特尔® 深度学习加速技术和英特尔® AMX

英特尔® SGX 最大飞地容量高达 64 GB

可通过英特尔® QAT、英特尔® DLB、英特尔® DSA 和英特尔® IAA 加速工作负载



- ¹ 第三代英特尔® 至强® 可扩展处理器与 AMD EPYC 处理器相比。详情请见以下网址的 [I26–I30]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/processors/> (第三代英特尔® 至强® 可扩展处理器)。结果可能不同。
- ² 详情请见以下网址的 [N9]: <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。
- ³ 详情请见以下网址的 [G1]: <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。
- ⁴ 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。
- ⁵ 详情请见以下网址的 [D1]: <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。
- ⁶ 详情请见以下网址的 [N18]: <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。
- ⁷ 详情请见以下网址的 [N16]: <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。
- ⁸ 详情请见以下网址的 [G2]: <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 第四代英特尔® 至强® 可扩展处理器每个 CPU 有 80 条 PCIe 5.0 通道, 并支持 Flex Bus/CXL, 而第三代英特尔® 至强® 可扩展处理器则是每个 CPU 有 64 条 PCIe 4.0 通道。
- ¹⁰ 第四代英特尔® 至强® 可扩展处理器具备 8 通道 DDR5, 速率高达 4,800 MT/s (1DPC), 而第三代英特尔® 至强® 可扩展处理器配备 8 通道 DDR4, 速率为 3,200 MT/s (2 DPC)。
- ¹¹ 仅部分英特尔® 至强® 金牌 6400 处理器提供多达四路的可扩展性。

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 www.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。

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。

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。

本文中的表述是关于未来计划或期望的前瞻性表述。该表述基于当前预期, 且存在许多风险和不确定因素可能造成实际结果与本表述明示或默示的结果有实质性不同。详见英特尔在美国证券交易委员会的备案文件, 可见于 www.intc.com。

© 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。