

英特尔® 至强® 6 处理器家族



英特尔® 至强® 6 处理器家族引入强大的计算平台，性能和能效表现双优，对满足现代数据中心不断变化的需求至关重要。从支持计算密集型 AI 到赋能可扩展云原生微服务，该处理器家族均可灵活满足不同的运营需求。

更好地满足当今数据中心的需求

数据中心基础设施是企业 and 机构最重要的投资之一。IT 负责人不仅需要利用这些资源来促进业务增长，还要权衡其安全性、能效、可管理性以及影响总体拥有成本 (TCO) 的其他因素。

也许最重要的是，IT 负责人必须要了解推动数据中心需求迅速变化的市场因素。例如，在工作负载自动化和生成式 AI (GenAI) 等技术变革的推动下，AI 的市场规模及其带来的机遇预计在未来五年内将增长 4 倍¹。越来越多的企业应用开始添加推理能力，这需要处理大型数据矢量，因此需要有并行数据计算能力以及更高单核性能的产品。此外，随着以云原生原则重新设计的单体应用面世，云微服务市场预计将增长 5 倍²。这些工作负载属于并行任务，因此处理这些任务更好的办法是借助高效的标量处理能力，而非增加复杂的计算能力。

通过加速多种工作负载更快获得业务成果

凭借更多的内核、灵活的微架构、更大的内存带宽和出色的输入/输出 (I/O)，英特尔® 至强® 6 处理器家族让多种工作负载的性能和能效表现再创新高。其全新功能和内置加速器为目标工作负载带来进一步助力，实现了更高的性能和能效。

英特尔® 至强® 6 处理器与第五代英特尔® 至强® 可扩展处理器比较



英特尔® 至强® 6 处理器与第二代英特尔® 至强® 可扩展处理器比较



出色的用户体验

英特尔® 至强® 6 处理器提供高水准的质量和可靠性，这是英特尔® 产品一贯为客户所称道的优势。保持持续运行并尽可能减少系统维护时间是管理数据中心服务级别协议 (SLA) 和整体 TCO 的基础。英特尔® 产品的可靠性、可用性和可维护性 (RAS) 是由一系列功能作为支撑的，这些功能可以帮助延长系统正常运行时间、减少计划外停机时间并保持数据完整性。

安全性对于用户体验和客户满意度非常重要。无论在本地还是云端，IT 团队都必须防范越来越多的安全威胁，并始终遵守隐私法规。为了保护使用中的数据，英特尔® 至强® 处理器采用了机密计算技术，能够更好地满足业务和监管要求。英特尔® Trust Domain Extensions (英特尔® TDX) 提供虚拟机 (VM) 层面的隔离和保密性，而英特尔® 软件防护扩展 (Intel® Software Guard Extensions, 英特尔® SGX) 则提供应用层面的隔离。

性能和能效不打折扣

英特尔® 至强® 6 处理器家族引入了创新的模块化 x86 架构，使数据中心架构师能够配置和部署专为企业特定需求和跨私有云、公共云和混合云工作负载而构建的基础设施。如表 1 所示，英特尔® 至强® 6 处理器有四个不同系列，提供不同的性能水平，可以满足从入门级到要求严苛的工作负载的需求。与前几代产品相比，这四个系列的内核数、缓存、内存速度和容量以及 I/O 均有提升。

为充分灵活满足不同需求，英特尔® 至强® 6 处理器采用两种不同的 CPU 微架构：P-core (性能核) 和 E-core (能效核)。两个版本均使用兼容的 x86 指令集架构 (ISA) 和通用硬件平台 (包括 CPU 插槽类型)。此外，英特尔还与行业合作伙伴合作，通过通用操作系统、编译器、库和框架来确保对两种内核类型的无缝使用。有了这种共享的软件堆栈以及由软硬件供应商组成的全球生态系统，解决方案便可以满足不同的业务需求。

表 1. 英特尔® 至强® 6 处理器家族包含四个系列处理器

系列	应用场景
英特尔® 至强® 6900 系列处理器	旗舰级，适合要求严苛的云、AI 和科学计算环境
英特尔® 至强® 6700 系列处理器	高端级，适合各种数据中心和电信环境
英特尔® 至强® 6500 系列处理器	主流级，适合主流服务器和边缘环境
英特尔® 至强® 6300 系列处理器	入门级，适合中小型企业环境

英特尔® 至强® 6 性能核处理器

经过精心优化，性能核的单核性能非常出色，拥有比其他通用 CPU 更好的 AI 性能，能够从容应对广泛的工作负载。与第五代英特尔® 至强® 可扩展处理器（常用于较新的计算密集型解决方案）相比，英特尔® 至强® 6 性能核处理器的性能提升高达 2 倍¹³。

- 每个内核均具备 AI 加速功能，更高效地助力实现 AI 无处不在。英特尔® 高级矩阵扩展（Intel® Advanced Matrix Extensions，英特尔® AMX）可加快 int8 和 BF16 的推理速度，并为以 FP16 精度训练的模型提供更好支持（int8 的速度高达 2,048 FLOPS，BF16/FP16 则高达 1,024 FLOPS）。
- 采用更快的 DDR5 内存和 Multiplexed Rank DIMM (MRDIMM)，内存吞吐量得以显著提高。与 RDIMM 相比，内存带宽提高了 37% 以上¹⁴，预计数据传输速率高达 8,800 MT/s。
- 每路多达 128 个内核，三级缓存高达 480 MB，能够以较低的时延从三级缓存访问大量数据。

高达
2 倍
性能提升

（与第五代英特尔® 至强® 可扩展处理器相比）¹³

MRDIMM
实现

37%
以上的
内存带宽提升

（与 RDIMMs 相比）¹⁴

英特尔® 至强® 6 能效核处理器

经过精心优化，能效核实现了更高的内核密度，拥有更出色的每瓦性能，为需要较高任务并行吞吐量的云级工作负载带来明显优势。与第二代英特尔® 至强® 可扩展处理器（目前大部分数据中心系统使用的处理器且每瓦性能出色）相比，英特尔® 至强® 6 能效核处理器的每瓦性能可实现 2.6 倍以上的提升¹²。这一出色能效表现使其在电力、空间和散热受限的情况下也是理想选择。英特尔® 至强® 6 能效核处理器可以：

- 将 4.3 个基于第二代英特尔® 至强® 可扩展处理器的服务器替换为 1 个性能相似的服务器¹⁵。
- 将 3 个基于第二代英特尔® 至强® 可扩展处理器的系统机架整合为 1 个机架¹⁶。
- 利用英特尔® 高级矢量扩展 2（Intel® Advanced Vector Extensions 2，英特尔® AVX2）和神经网络指令（VNNI）以及将精度快速转换成 BF16 和 FP16 等全新增强功能，加速 AI 推理和面向矢量的运算。
- 每路多达 288 个内核，三级缓存高达 216 MB，能够以较低的时延从三级缓存访问大量数据。

可将
3 个
机架
整合为 1 个¹⁶

实现

2.6 倍
以上的
每瓦性能提升

（与第二代英特尔® 至强® 可扩展处理器相比）¹²

英特尔® 至强® 6 性能核处理器和英特尔® 至强® 6 能效核处理器之所以能效出众，是因为它们能够随着服务器占用率的提高而提供可扩展的每瓦性能，使整个负载线的功耗/性能呈近线性关系。对于性能要求严苛的工作负载，这意味着在高负载下可以高效利用电能，以更快地完成任务。对于云或共享计算环境中常用的可扩展实施方案，这种能效水平意味着服务器在欠载时仅消耗所需的电能，从而在实例未被充分利用时降低成本。

英特尔® 至强® 6 处理器还支持全系统电源管理和遥测功能，可持续性得到了进一步增强。这些功能可以提高每个应用的每瓦性能，有助于降低总体能耗。

英特尔® 至强® 6 处理器的广泛用途和互补优势

针对需求截然不同甚至极端的工作负载，英特尔® 至强® 6 性能核处理器是 AI 等计算密集型、基于矢量的工作负载的理想之选；而英特尔® 至强® 6 能效核处理器则适合微服务等任务并行、基于标量的工作负载。对于需求介于这两种极端情况之间的工作负载，这两种微架构相结合可让解决方案互补且用途更广泛。例如，可使用搭载英特尔® 至强® 6 能效核处理器的系统来节省电能，以便为搭载英特尔® 至强® 6 性能核处理器的节点提供足够的电能来运行 AI 和科学工作负载。采用英特尔® 至强® 6 性能核处理器和英特尔® 至强® 6 能效核处理器混合设计的数据中心可以利用其平台通用性，根据性能和功耗需求将工作负载从一种内核的处理器转移到另一种内核的处理器。这种丰富的选项组合使数据中心可以随着业务增长轻松扩展。

英特尔® 至强® 6 处理器的互补优势还在于，数据中心可以轻松地混合使用基于英特尔® 至强® 6 性能核处理器和英特尔® 至强® 6 能效核处理器的服务器，以支持需要不同结构的数据库的业务需求。对于涉及复杂的数据关系、复杂的查询、连接和聚合的关系型数据库，就可以利用英特尔® 至强® 6 性能核处理器的并行数据处理能力。涉及大量小型独立数据检索请求（如键值存储）的非关系型数据库，则可以利用英特尔® 至强® 6 能效核处理器的任务并行设计。

亮点技术

英特尔® 至强® 6 处理器家族创新的性能核和能效核微架构具有以下先进功能和优势：

- 单路英特尔® 至强® 6 能效核处理器拥有多达 288 个内核，单路英特尔® 至强® 6 性能核处理器拥有多达 128 个内核，实现了更高密度计算性能和可扩展性。
- 对于基于 BF16 和 FP16 的模型，英特尔® AMX 的乘法累加 (MAC) 运算速度比英特尔® 高级矢量扩展 512 (Intel® Advanced Vector Extensions 512, 英特尔® AVX-512) 提升高达 16 倍，AI 性能显著增强（英特尔® 至强® 6 性能核处理器的功能）。
- 英特尔® AVX-512 包含特有的指令，每个内核拥有两个 512 位融合乘加 (FMA) 单元，大幅提高了 AI、科学计算和数据库工作负载常见的矢量计算速度（英特尔® 至强® 6 性能核处理器的功能）。
- 支持 VNNI 指令的英特尔® AVX2 以及将精度快速转换为 BF16 和 FP16 的能力为英特尔® 至强® 6 能效核处理器提供了更好的 AI 兼容性。
- 与标准 DDR5 DIMM 相比，MRDIMM 能够提供超过 37% 的额外内存带宽，可支持 AI 和科学计算中的带宽受限用例（英特尔® 至强® 6 性能核处理器的功能）¹⁴。
- 多达 12 条内存通道，进一步支持更高的内存带宽。
- 英特尔® 超级通道互联 (Intel® Ultra Path Interconnect, 英特尔® UPI) 2.0 的跨插槽内带宽速度高达 24 GT/s，与上一代产品相比提升高达 20%。
- 双路服务器拥有多达 188 条 PCIe Gen 5 通道，单路服务器则多达 136 条，可以支持重要的 I/O 附加组件，包括加速器、网络适配器、存储控制器和存储。
- 多达 64 条 Compute Express Link (CXL) 2.0 通道，每条通道的数据传输速率高达 32 GT/s，支持 CXL 功能，包括内存扩展和共享（包括 Type 3 设备）。
- 当使用低成本内存（如支持 CXL 2.0 的 DDR4）时，“Flat”内存模式可帮助扩展系统内存并优化 TCO。
- 英特尔® 数据保护与压缩加速技术 (Intel® QuickAssist Technology, 英特尔® QAT) 支持卸载批量加密和压缩，以加速网络和存储。
- 英特尔® 数据流加速器 (Intel® Data Streaming Accelerator, 英特尔® DSA) 2.0 能够卸载数据传输和转换操作，例如移动、填充、比较、循环冗余校验 (CRC)、数据完整性字段 (DIF)、增量和刷新。
- 英特尔® 存内分析加速器 (Intel® In-Memory Analytics Accelerator, 英特尔® IAA) 可以卸载内存压缩和解压缩、扫描和过滤功能以及循环冗余校验。
- 英特尔® 动态负载均衡器 (Intel® Dynamic Load Balancer, 英特尔® DLB) 支持动态分配网络数据包处理和卸载重排序操作。
- 英特尔® TDX 用 AES-256 和 2,048 个加密密钥进行了升级，机密计算能力进一步增强，能够更好地保护敏感的企业数据。
- 英特尔® On Demand 服务使硬件提供商可以启用部分基于 CPU 的特性和功能。它通过两种模式提供服务：基于一次性许可证激活功能，以及基于用量付费。

进一步了解英特尔® 至强® 6 处理器以及上文列举的功能特性，请访问 intel.cn/xeon。

英特尔® 至强® 6 处理器家族概览

英特尔® 至强® 6900 系列处理器采用全新的英特尔® 服务器平台设计，可为客户提供更出色性能、更高内存带宽和更高吞吐量，非常适合云、科学计算和 AI 环境。该系列处理器拥有更多内核、更多的内存通道和 I/O 通道，热设计点高于其他系列。

英特尔® 至强® 6700 系列和英特尔® 至强® 6500 系列处理器则采用经过升级的服务器平台设计，具备出色的性能、成本和能效优势，非常适合广泛的数据中心环境。这些处理器支持单路至 8 路配置，并可在既有的数据中心供电和散热条件下提供更大的 I/O 和内存。

英特尔® 至强® 6 处理器

英特尔® 至强® 6900 系列 处理器

旗舰级

采用全新的英特尔® 服务器平台设计，非常适合云计算、AI、科学计算、软件即服务 (SaaS) 和基础设施即服务 (IaaS) 等工作负载。

英特尔® 至强® 6 性能核 CPU

- 每个 CPU 拥有多达 128 个内核 (256 个线程)
- 每个 CPU 高达 500 W
- 单路或双路服务器
- 12 条内存通道
- 高达 6,400 MT/s DDR5
- 8,800 MT/s MRDIMM
- 多达 96 条 PCIe 5.0 通道
- 6 条英特尔® UPI 2.0 链路

即将推出

英特尔® 至强® 6 能效核 CPU

- 每个 CPU 拥有多达 288 个内核 (288 个线程)
- 每个 CPU 高达 500 W
- 单路或双路服务器
- 12 条内存通道
- 高达 6,400 MT/s DDR5
- 多达 96 条 PCIe 5.0 通道
- 6 条英特尔® UPI 2.0 链路

即将推出

英特尔® 至强® 6500 系列/ 6700 系列 处理器

高端级

对现有的英特尔® 服务器平台进行了大幅升级。面向企业 IT、数字服务提供商和电信的主流边缘协同服务器。非常适合 AI、科学计算、网络和媒体、数字服务、基础设施和存储、Web、应用以及微服务等工作负载。

- 每个 CPU 拥有多达 86 个内核 (172 个线程)
- 每个 CPU 高达 350 W
- 单路、双路、四路或八路服务器
- 8 条内存通道
- 高达 6,400 MT/s DDR5
- 高达 8,000 MT/s MRDIMM
- 多达 88 条 PCIe 5.0 通道，其中单路设计最高可达 136 条
- 4 条英特尔® UPI 2.0 链路

即将推出

- 每个 CPU 拥有多达 144 个内核 (144 个线程)
- 每个 CPU 高达 330 W
- 单路或双路服务器
- 8 条内存通道
- 高达 6,400 MT/s DDR5
- 多达 88 条 PCIe 5.0 通道
- 4 条英特尔® UPI 2.0 链路

现已上市



¹ Fortune Business Insights, "Artificial Intelligence Market Size, Share & Industry Analysis... (AI 市场规模、份额和行业分析...)", 2024 年 4 月。

fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114。

² Fortune Business Insights, "Cloud Microservices Market Size, Share & COVID-19 Impact Analysis... (云微服务的市场规模、份额与新冠疫情影响分析...)", 2024 年 4 月。

fortunebusinessinsights.com/cloud-microservices-market-107793。

³ 详情请见以下网址的 [9G10]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

⁴ 详情请见以下网址的 [9A10]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

⁵ 详情请见以下网址的 [9H10]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

⁶ 详情请见以下网址的 [7W5]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

⁷ 详情请见以下网址的 [7D2]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

⁸ 详情请见以下网址的 [7W2]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

⁹ 详情请见以下网址的 [7W4]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

¹⁰ 详情请见以下网址的 [7D1]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

¹¹ 详情请见以下网址的 [7W1]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

¹² 详情请见以下网址的 [7N1]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

¹³ 基于英特尔截至 2023 年 8 月的架构预测, 与第五代英特尔® 至强® 可扩展处理器相比性能更优。

¹⁴ 与 DDR5 6,400 RDIMM 比较。

¹⁵ 基于特定 SLA, MySQL OLTP 和服务端 Java 吞吐量。详情请见 intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

¹⁶ 详情请见以下网址的 [7T1] 和 [7T2]: intel.com/processorclaims (英特尔® 至强® 6 处理器)。结果可能不同。

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。

© 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。