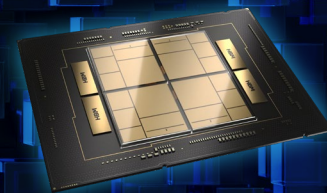


产品简介

计算加速
系统与图形

intel®

英特尔® 至强® CPU Max 系列



过去十年，随着人工智能 (AI) 技术的加入，峰值算力大幅增长，但由于在向内核传输数据时效率低，因此工作负载性能未能同步提升。英特尔® 至强® CPU Max 系列的诞生，使英特尔® 至强® 平台如虎添翼，它是英特尔唯一一个基于 x86 架构并采用高带宽内存的 CPU 系列，可释放和加速内存密集型科学计算和 AI 工作负载。



从天气预报到绘制人类基因组图谱和助力治疗全球范围内的致命疾病，再到设计更节能的材料，科学计算触及到人类生活的方方面面。虽然科学计算和 AI 领域的进步提高了竞争力并将对科学计算的需求带至新的高度，但却没有“放之四海而皆准”的解决方案。传统的科学计算软件种类繁多，如果依据垂直领域和特性对常见的工作负载进行划分，有些属于内存密集型，有些则属于计算密集型；有些需要有大量控制流的小型内核 (kernel)，有些则需要支持数据并行处理的大型内核；但大多都涉及非常大的数据集。

英特尔® 至强® CPU Max 系列采用高带宽内存 (HBM)，为英特尔® 至强® 可扩展处理器家族注入新力量，它专为建模、人工智能、深度学习、科学计算和数据分析等数据密集型工作负载设计，专注解锁性能，加速新发现。

更高带宽，更优性能

英特尔® 至强® CPU Max 系列采用全新微架构，支持一系列可提升平台能力的特性，包括更多内核、先进的 I/O 与内存子系统，以及可加速重大发现的内置加速器。英特尔® 至强® CPU Max 系列具有以下特性：

- 多达 56 个 P-core (性能核)：内核由 4 个小芯片构成，采用英特尔的嵌入式多芯片互连桥接 (EMIB) 技术连接，功耗为 350 W。
- 64 GB 高带宽封装内存及 PCIe 5.0 和 CXL 1.1 I/O。英特尔® 至强® CPU Max 系列每核均具备高带宽内存 (HBM) 容量，可满足大多数常见科学计算工作负载的要求。
- 与其他 CPU 相比，在使用 Numenta 的 AI 技术进行自然语言处理 (NLP) 时，其高带宽内存 (HBM) 优势可带来高达 20 倍的性能提升²。

5 倍 性能提升

(与第三代英特尔® 至强® 可扩展处理器在内存带宽方面的对比)¹

双路英特尔® 至强® CPU MAX 系列处理器与双路英特尔® 至强® 铂金 8380 处理器相比

加速科学创新

支持快速获得发现、高效开展科研活动。使用英特尔® 至强® CPU Max 系列和第四代英特尔® 至强® 可扩展处理器，您不但可以获得处理要求严苛的工作负载所需的性能与能效，还可得到各种出色的内置加速器的助力。您可以利用面向科学计算和 AI 工作负载的关键加速器，提高 CPU 使用效率、降低功耗、实现更高的投资回报率 (ROI)：

- **英特尔® 高级矩阵扩展 (Intel® Advanced Matrix Extensions, 英特尔® AMX) :** 英特尔® AMX 可显著加速基于 CPU 的深度学习推理和训练，从而进一步提升 AI 性能，与英特尔® 高级矢量扩展 512 (Intel® Advanced Vector Extensions 512, 英特尔® AVX-512) 相比，它在进行 INT8/INT32 累加运算时可将峰值吞吐量提高 8 倍³。
- **英特尔® 数据流加速器 (Intel® Data Streaming Accelerator, 英特尔® DSA) :** 可通过加速流数据传输，提升数据密集型工作负载的性能。在使用 NVMe over TCP 时，英特尔® DSA 可以将存储 IOPS 提升高达 79%，将时延降低多达 45%⁴。
- **英特尔® 高级矢量扩展 512 (Intel® Advanced Vector Extensions 512, 英特尔® AVX-512) :** 通过矢量化实现性能提升，使更大数据集上的运算速度更快，满足科学仿真、AI/深度学习、3D 建模与分析以及其他密集型工作负载的要求。英特尔® AVX-512 是最新的 x86 矢量指令集，可为要求严苛的计算任务提升性能。
- **I/O 和内存子系统升级包括：**
 - **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⁴。
 - **PCI Express Gen 5 (PCIe 5.0):** 带来全新的 I/O 速度，可在 CPU 和设备之间实现更高的吞吐量。第四代英特尔® 至强® 可扩展处理器和英特尔® 至强® CPU Max 系列处理器配备多达 80 条 PCIe 5.0 通道，其 I/O 带宽是 PCIe 4.0 的两倍⁴。
 - **Compute Express Link (CXL) 1.1:** 提供高网络带宽支持，并使附加加速器高效运行。
- **与英特尔® 至强® 平台实现轻松整合：** 由于处理器插槽 (Socket) 配置相同，可轻松将英特尔® 至强® CPU Max 系列处理器添加到第四代英特尔® 至强® 可扩展平台，并且在大多数部署方案中都无需更改代码。

灵活应对各种科学计算和 AI 工作负载

英特尔® 至强® CPU Max 系列处理器具备出色的灵活性，可根据工作负载的特性，在不同的内存模式或配置下运行：

- **“仅 HBM”模式：** 该模式支持内存容量需求不超过 64 GB 的工作负载以及每核 1 至 2 GB 的内存扩展能力，同时无需更改代码和另购 DDR，即可启动系统。
- **“HBM Flat”模式：** 该模式可为需要大内存容量的应用提供灵活性，它通过 HBM 和 DRAM 提供一个平面内存区域 (flat memory region)，适用于每核内存需求大于 2 GB 的工作负载。使用该模式时可能需要更改代码。
- **“HBM 缓存”模式：** 旨在提升内存容量需求大于 64 GB 或每核内存需求大于 2 GB 的工作负载的性能。使用该模式时，无需更改代码，且 HBM 可缓存来自 DDR 的事务。

英特尔® 至强® CPU Max 系列	
内核数	32-56
HBM2e 内存	64 GB
HBM 最大传输速率	3200 MT/s
DDR5 最大传输速率	4800 MT/s (1 个 DPC) 4400 MT/s (2 个 DPC)
加速器	AMX, 4 个英特尔® DSA
AI/ML 指令	INT8 和 BFLOAT16

跨多架构加速科学计算和 AI 工作负载

整个英特尔® 至强® CPU Max 系列的产品均得到 oneAPI 的支持。oneAPI 是一个统一的、基于标准的开放式通用编程模型，可释放生产力并解锁性能。开发人员可利用英特尔® oneAPI 工具套件以及面向特定领域的专用工具套件，打造跨多种架构运行的通用计算、科学计算和 AI 应用，并对其进行分析、优化和扩展。这些资源包括矢量化、多线程、多节点并行和内存优化方面的前沿技术，让您轻松构建随时能为科学计算所用的高性能、多架构软件。如需了解最新科学计算软件开发人员工具，请访问[适用于第四代英特尔® 至强® 可扩展处理器及英特尔® 至强® CPU Max 系列处理器的软件和科学计算软件和工具资源页面](#)。



英特尔® 至强® CPU Max 系列处理器

SKU 编号	内核数	基频 (GHz)	全核睿频 (GHz)	最大睿频 (GHz)	缓存 (MB)	TDP (W)	最大可扩展性	DDR5 内存速度	英特尔® SGX 飞地容量 (每个处理器)
9480	56	1.9	2.6	3.5	112.5	350	双路	4800	512 GB
9470	52	2	2.7	3.5	105	350	双路	4800	512 GB
9468	48	2.1	2.6	3.5	105	350	双路	4800	512 GB
9460	40	2.2	2.7	3.5	97.5	350	双路	4800	128 GB
9462	32	2.7	3.1	3.5	75	350	双路	4800	128 GB

如需了解最新信息，请访问

intel.cn/content/www/cn/zh/products/details/processors/xeon/max-series.html



¹ 配合工作负载/配置信息请见 intel.cn/performanceindex (前往 Events 项下的 Supercomputing 22)。结果可能不同。

² Numenta BERT-Large

- 英特尔® 至强® 铂金 8480+ 处理器: 基于 2022 年 11 月 28 日 Numenta 所做的测试。单节点, 2 个英特尔® 至强® 铂金 8480+ 处理器, 512 GB DDR5-4800, Ubuntu 22.04 内核 5.17, OpenVINO 2022.3, 经 Numenta 优化的 BERT-Large, 序列长度为 512, 批大小为 1
- 英特尔® 至强® CPU Max 9468 处理器: 基于 2022 年 11 月 30 日 Numenta 所做的测试。单节点, 2 个英特尔® 至强® CPU Max 9468 处理器, 128 GB HBM2e@3200 MT/s, Ubuntu 22.04 内核 5.15, OpenVINO 2022.3, 经 Numenta 优化的 BERT-Large, 序列长度为 512, 批大小为 1

³ 来源: 英特尔网站 P-core (性能核) 声明和细节部分有关英特尔® AMX 的描述。“性能指标: 2021 年英特尔架构日”, edc.intel.com/content/www/cn/zh/products/performance/benchmarks/ (前往 Events 项下的 Architecture Day 2021)。

⁴ 更多信息请见 www.intel.cn/PerformanceIndex。实际性能受使用情况、配置和其他因素的差异影响。结果可能不同。

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。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。英特尔技术可能需要启用硬件、软件或激活服务。

© 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。