

第三代英特尔®至强®可扩展处理器

在以数据为中心的时代助力转型

在不断演进的数字世界,能够带来突破或颠覆的新兴技术层出不穷,这一趋势对全球经济的影响也在日益强化。随着我们步入"新十年",众多公司将越来越看重创建数字化增强型产品、服务和体验的能力,而大型企业和机构则会借力于技术进步及随之而来的创新应用模式。

这一全球范围的转型,迅速激发了对于更灵活的计算、网络和存储的需求。能够无缝扩展以实现即时响应和支持多样化性能要求的基础设施,对未来的工作负载至关重要。数据的生成和使用呈指数级增长、云上的大规模计算和 5G 网络的迅速扩展,以及高性能计算 (HPC) 和人工智能 (AI) 在全新应用场景中的融合,无不要求当今的数据中心和网络与时俱进,否则就会在竞争激烈的环境中落伍。在这些需求的推动下,面向未来的现代化数据中心和网络的架构必须能够迅速而又灵活地扩展。

英特尔®至强®可扩展平台则为数据中心的敏捷性和可扩展性实现飞跃奠定了基础。这款创新型处理器采用突破性的设计,在计算、存储、内存、网络和安全方面,达到了全新的平台级融合水平。企业、云服务提供商和通信服务提供商现在可以借助一个功能丰富、高度灵活的平台,来推进其雄心勃勃的数字转型计划。



更强性能强化数据洞察力

第三代英特尔®至强®可扩展处理器是一款集成了人工智能加速技术,并针对工作负载进行了优化的平台,为企业在多云环境到智能边缘之间充分利用数据奠定了无缝的性能基础。

- 更出色的性能:提供更高的多路内核密度:每颗处理器 支持的内核数多达 28 个;在 8 路配置中,每个平台支持 的内核数量可多达 224 个。与第二代英特尔®至强®可扩展 处理器相比,性能、吞吐量和 CPU 频率均得到了提升。
- 支持 VNNI 和全新 bfloat16 加速的增强型英特尔[®] 深度 学习加速技术(英特尔[®] DL Boost): 增强版英特尔[®] 深度学习加速技术是首款同时支持 16 位 Brain Floating Point (blfoat16) 和矢量神经网络指令 (VNNI) 的 x86 技术,增强了人工智能推理和训练性能。与上一代产品 相比,人工智能训练性能提升高达 1.93 倍¹。
- 更多英特尔[®] 超级通道互联(英特尔[®] UPI): 与上一代产品相比,第三代英特尔[®] 至强[®] 可扩展处理器支持多达6个英特尔[®] UPI通道,增强了平台可扩展性并提升了面向I/O 密集型工作负载的 CPU 之间的带宽,在提升吞吐量和能效之间实现了更好的平衡。

- **DDR4** 内存速度和容量提升: 内存子系统方面的增强包括 支持多达 6 个 DDR4-3200 MT/s 的通道和 16 Gb DIMM, 每路支持高达 256 GB DDR4 DIMM。
- 英特尔[®] 高级矢量扩展 512 技术(英特尔[®] AVX-512): 在建模和仿真、数据分析和机器学习、数据压缩、可视化、 以及数字内容创作等应用领域,英特尔[®] AVX-512 技术 能够为要求苛刻的计算任务提升性能和吞吐量。
- 支持英特尔® 傲腾™ 固态盘和英特尔® QLC 3D NAND 固态盘: 具备高吞吐量、低延迟、高服务品质 (QoS) 和高耐用性等优势,助力突破数据访问瓶颈。全新的英特尔® 固态盘 D7-P5500 和 D7-P5600 融合了低延迟 NVMe* 互连、英特尔的低开销存储软件以及先进的 TLC 3D NAND,专为数据中心和边缘上部署的人工智能和数据分析工作负载的密集 I/O 需求而打造。



业务弹性

第三代英特尔®至强®可扩展处理器集成了硬件增强型安全特性,更有助于在维持工作负载完整性和降低性能开销的同时阻止恶意攻击。在空闲、使用中和忙碌三种状态下,都能以更高的可用性和更高的加密效率实现更为可信的服务交付。



• 新型英特尔[®] Platform Firmware Resilienc (英特尔[®] PFR): 这是一个基于英特尔[®] FPGA 的解决方案,可以更好地保护平台固件、检测损坏,并恢复到已知良好 (known-good) 状态³。



• 英特尔[®] Security Essentials 和英特尔[®] Security Libraries for Data Center: 支持通过硬件增强型信任根构建更可信、更安全、也更加可控的云³。



• 集成英特尔[®] QuickAssist 技术(英特尔[®] QAT): 运用基于芯片组的 硬件加速能力来处理日益增多的压缩和加密工作负载,从而提高效率, 同时在服务器、存储和网络基础设施之间提供高达 100 GB/s 的增强型 数据传输和保护能力³。 特定款英特尔[®] C620A 系列芯片组支持此项 技术。



敏捷服务交付

第三代英特尔[®] 至强[®] 可扩展平台为计算、网络、存储和持久内存带来了创新和硬件增强型虚拟化的支持。它能更好地助力在经济高效、灵活而又可扩展的多云环境中持续打造出色的企业对企业 (B2B) 和企业对消费者 (B2C) 体验。

- 英特尔® Speed Select 技术: 英特尔® Speed Select 技术 (英特尔® SST)汇集众多可配置的处理器功能,实现了 处理资源的优化,以增强工作负载性能、提高利用率并 有助于优化平台总体拥有成本。特定款第三代英特尔® 至强® 可扩展处理器支持此项技术。
- 支持英特尔[®] 傲腾™ 持久内存 200 系列(英特尔[®] 傲腾™ PMem): 与 DRAM 相比, 英特尔[®] 傲腾™ 持久内存 200 系列具有更优的性价比, 可辅以 DRAM 实现更大的系统内存容量, 从而加速工作负载处理和服务交付。四路配置的第三代英特尔[®] 至强[®] 可扩展平台设计支持的英特尔[®] 傲腾™ 持久内存 200 系列模组容量高达 512 GB, 每路可提供多达 3 TB 持久内存。
 - 英特尔® 傲腾™ 持久内存实现系统内存容量突破,支持每路处理器高达 4.5 TB 的内存容量,在与 DRAM 结合使用时,可实现总计高达 18 TB 的系统内存。
 - 与第一代产品相比,英特尔[®] 傲腾™ 持久内存 200 系列 提供的内存带宽平均高出 25%⁴。

- 英特尔® Infrastructure Management 技术(英特尔® IMT): 英特尔® Infrastructure Management 技术是一种资源管理框架,它包含具备平台级检测、报告和配置功能的英特尔® 资源调配技术(英特尔® RDT)和英特尔® 虚拟化技术(英特尔® VT-x)。对资源进行硬件增强型的监控、管理和控制,有助于提高数据中心资源的效率、利用率和安全性。
- 应用设备队列 (ADQ): 这是英特尔®以太网 800 系列网络适配器中的一项开放型技术,由于这种队列中没有其他应用共享带宽,因而可为特定应用提供无竞争且更流畅的数据传输支持。
- 英特尔® Stratix® 10 NX FPGA: 英特尔® Stratix® 10 NX FPGA 专为需要实时跨多节点评估多个变量的人工智能应用而打造,能够为自然语言处理和金融欺诈检测等工作负载提供更出色的性能。这种 FPGA 十分灵活,它通过一体封装的高带宽内存,加速矩阵和矢量运算以及集成的高吞吐量网络,来提供更优的人工智能算力。此外,此款FPGA 产品还能随着模型的变化和扩展重新进行编程和优化。

运用英特尔® 精选解决方案加速价值实现进程

在当今越趋纷繁复杂的数据中心中,软硬件基础设施不可"一刀切"。英特尔[®] 精选解决方案针对实际性能进行了优化,并经过了严格的基准测试和验证,尽可能消除各种不确定性因素。这些解决方案可加速在英特尔[®] 至强[®] 可扩展处理器上的基础设施部署,可满足当今高级数据分析、混合云、存储和网络通信中的关键工作负载的需求。

英特尔[®] 精选解决方案由解决方案合作伙伴提供,英特尔按照严格的质量、性能和安全标准进行了验证。

如需了解搭载第三代英特尔®至强®可扩展处理器的英特尔® 精选解决方案,请访问 intel.com/selectsolutions



支持 bfloat16 加速的英特尔® 深度学习加速技术,让集成人工智能加速的处理器的性能更上一层楼

创新型算法、数据来源和体量上的新变化,以及计算和存储领域的技术进步,都在驱动着当今科学发现的演进。机器 学习、深度学习和人工智能将大量计算功能与海量数据融合在一起,推进了自主系统和自动驾驶汽车等下一代应用的 发展。

第三代英特尔® 至强® 可扩展处理器专为灵活运行的目标而设计,让企业能够在与既有工作负载相同的硬件平台上运行复杂的人工智能工作负载。增强型英特尔® 深度学习加速技术是首款同时支持 16 位 Brain Floating Point (blfoat16)数字格式和矢量神经网络指令 (VNNI)的 x86 技术,增强了人工智能推理和训练性能。与上一代产品相比,人工智能训练性能提升高达 1.93 倍,图像分类的推理性能提升达 1.87 倍^{1,5}。新增的 bfloat16 加速能力对医疗保健、金融服务和零售行业中诸如视觉、自然语言处理 (NLP) 和强化学习 (RL)等需要兼顾吞吐量和准确率的应用大有裨益。与上一代相比,支持bfloat16 加速的英特尔® 深度学习加速技术为自然语言处理提供了高达 1.7 倍的人工智能训练性能®。第三代英特尔® 至强®可扩展处理器有助于从数据中心到边缘都实现人工智能就绪。

细分市场



企业与政府



云服务提供商



医疗保健



金融服务



雬隹

人工智能创新亮点

- 英特尔® 至强® 铂金 8300 处理器
- 英特尔® 傲腾™ 持久内存 200 系列
- 增强了 DDR 性能, 速度高达 3200 MT/s
- 英特尔® 深度学习加速技术,现已支持 bfloat16 加速
- 英特尔® 超级通道互联比上一代更多
- 英特尔® 高级矢量扩展 512 技术
- 英特尔® Omni-Path 架构 Host Fabric 接口
- 英特尔® 傲腾™ 固态盘

如需了解有关英特尔®至强®可扩展处理器集成的人工智能性能和加速技术的更多信息,请访问 ai.intel.com

面向图像分类的人工智能推理 性能提升高达 1.87 倍

与上一代产品相比5

面向自然语言处理的人工智能训练 性能提升高达 1.7 倍

与上一代产品相比6

面向自然语言处理的人工智能推理 性能提升高达 1.9 倍

与上一代产品相比7

为云、数据分析和关键业务型应用增强每个节点的性能

云行业和企业级客户渴望能够从激增的数据流中提取有价值的信息,快速获取有助于制定业务举措的洞察。企业使用的 传统和新兴应用(包括预测性分析、机器学习、高性能计算和关键业务型应用)需要全新水平的算力和大量分层的数据 存储容量。现代化数据中心在设计时采用了全面融合的方法。这种方法不仅能够更加灵活地提供新服务以及改善当前基础 设施资产的总体拥有成本,还能为自治的、混合的数据中心提供无缝程度和可扩展性都更优的支持。

英特尔® 至强® 可扩展平台是一种面向未来的平台,它能为企业提供多种全新的能力,既可以服务于混合云和数据驱动的时代,同时也有助于改善日常运营。这一多功能平台为计算密集型和延迟敏感型应用带来了突破性的计算性能以及内存和 I/O 领域的创新成果。通过与创新的英特尔® 傲腾™ 固态盘和英特尔® QLC 3D NAND 固态盘相结合,以英特尔® 至强® 可扩展平台为基础打造的平台可更高效地管理存储、缓存和内存间的海量数据,从而更轻松地满足数据和云时代的严苛需求。

细分市场



企业与政府



云服务提供商



制造



生命科学



油气



金融服务

云创新亮点

- 英特尔® 至强® 金牌 6300 和 5300 处理器
- 英特尔® 傲腾™ 持久内存 200 系列
- 英特尔® 超级通道互联比上一代更多
- 英特尔® 深度学习加速技术, 现已支持 bfloat16 加速
- 在八路配置中,每个节点内核数可多达 224
- 英特尔® 高级矢量扩展 512 技术
- 英特尔® Omni-Path 架构 Host Fabric 接口
- 英特尔® 傲腾™ 固态盘

如需了解第三代英特尔®至强®可扩展处理器为云、数据分析和关键业务型应用带来的优势,请访问 intel.com/cloud 每分钟处理的 OLTP 数据库事务 数量提升高达 1.98 倍 与 5 年前的四路平台相比⁸

6

采用可扩展式设计、提升每个节点的虚拟机密度

数字化服务增长迅猛,公有和私有的数据中心对增加虚拟机 (VM) 数量的需求也随之高涨。为提高数据中心的容量和效率, IT 部门总在想方设法提高虚拟机的性能和密度,每台服务器上加载的虚拟机正日趋增多。

第三代英特尔® 至强® 可扩展平台采用专为提升每个节点的虚拟机密度,并能实现在五代英特尔® 至强® 处理器和平台之间无缝迁移虚拟机而打造的处理器。这种多功能平台提升了处理器、内存、存储和 I/O 等资源的利用率,同时通过故障转移和恢复解决方案满足了服务水平协议的要求。企业可以利用英特尔® 至强® 可扩展平台将更多的应用整合到性能更高的服务器上,来优化空间、功耗、散热和维护成本,从而达到改善总体拥有成本的目的。在基于英特尔® 至强® 可扩展平台构建的基础设施上,搭配使用英特尔® 傲腾™ 持久内存 200 系列,也有助于企业快速预置虚拟机,改善基础设施的敏捷性和可扩展性,从而提高负载均衡、峰值工作负载管理、测试开发和系统维护的配置灵活性。





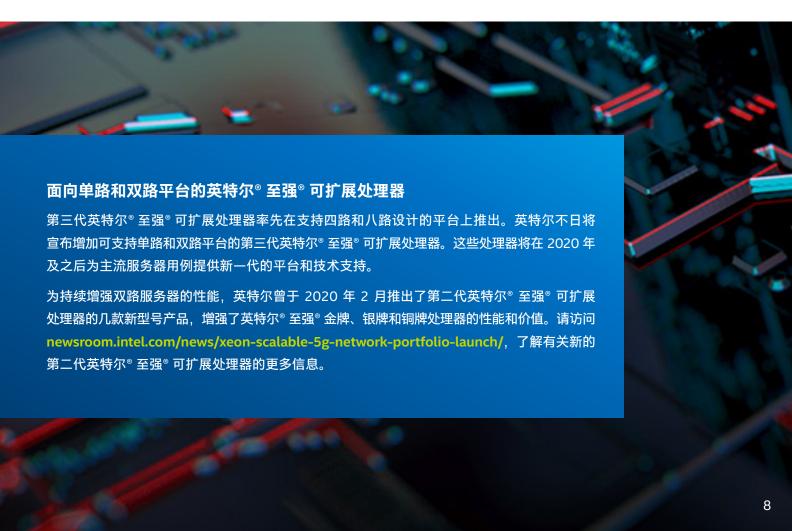


第三代英特尔®至强®可扩展处理器概览

英特尔® 至强® 铂金 8300 处理器是打造可靠、敏捷的混合云数据中心的坚实基础。这些处理器提供了硬件增强型安全特性,具有更出色的多路处理性能,专为关键业务型、实时分析、机器学习、人工智能和多云等工作负载而打造³。通过更加可信的硬件增强型数据服务交付,这些处理器为 I/O、内存、存储和网络技术均带来了改进,助力企业在数据越来越重要的世界中更高效地挖掘数据中蕴藏的价值和获得具有实践意义的数据洞察。全新的英特尔® 至强® 铂金 8300 处理器专为高级分析、人工智能和高密度基础设施而设计,它能提供更出色的性能、平台功能和工作负载加速能力。

- 每个英特尔® 至强® 可扩展处理器支持多达 28 个内核
- 每个处理器支持多达 6 个内存通道, 速度高达 3200 MT/s (1 DPC)
- 该处理器集成了支持 bfloat16 和 VNNI 的英特尔®深度学习加速技术,进一步提升了人工智能推理加速性能

英特尔[®] 至强[®] 金牌 6300 和 5300 处理器支持更高的内存速度、更大的内存容量以及多达四路的可扩展性,能够提供更强的性能、更大的内存容量、硬件增强型安全特性和工作负载加速功能。这些处理器针对要求严苛的主流数据中心、多云计算、网络和存储工作负载进行了优化。提供四路可扩展性,适合于范围更广的工作负载。



工作负载指南

第三代英特尔®至强®可扩展处理器适用于多种应用场景。该处理器具备的内核数、基频和睿频、内存配置和工作负载加速能力(如英特尔®深度学习加速和英特尔®Speed Select 技术)等特性可为特定工作负载和服务提供更强的性能支撑。本指南旨在帮助您探索第三代英特尔®至强®可扩展处理器。

	英特尔 [®] 深度 学习加速技术 (人工智能 训练)	英特尔 [®] 深度 学习加速技术 (人工智能 推理)	内存数据库	虚拟机密度	延迟敏感/ 高频	laaS/数据库 软件成本优化/ 通用计算	大型内存 多达4.5TB/路	英特尔 [®] Speed Select 技术
英特尔® 至强® 铂金 8380H 处理器	•							
英特尔® 至强® 铂金 8380HL 处理器			•				•	
英特尔® 至强® 铂金 8376H 处理器	•							
英特尔® 至强® 铂金 8376HL 处理器			•				•	
英特尔® 至强® 铂金 8354H 处理器	•	•						
英特尔®至强®铂金 8353H处理器								
英特尔® 至强® 金牌 6348H 处理器				•				
英特尔®至强®金牌 6328H 处理器					•			•
英特尔® 至强® 金牌 6328HL 处理器							•	
英特尔®至强®金牌 5320H 处理器						•		•
英特尔 [®] 至强 [®] 金牌 5318H 处理器						•		

全新英特尔® Speed Select 技术(英特尔® SST)功能

英特尔® Speed Select 技术汇集了众多功能,助您对处理器性能进行更精细的控制,从而优化总体拥有成本。有了英特尔® SST,服务器就能事半功倍。还能获得其他诸多效益。

英特尔® Speed Select Technology – Core Power (英特尔® SST-CP)

使工作负载能够因部分处理器内核的基频提升而变得更为 灵活。尽管各内核的最大睿频保持不变,但可以指定部分 内核以更高基频运行,而其他内核以较低基频运行。

英特尔[®] Speed Select Technology – Turbo Frequency (英特尔[®] SST-TF)

使工作负载能够因部分处理器内核的睿频提升而变得更为 灵活。尽管各内核的基频保持不变,但可以指定部分内核 以更高睿频运行,而其他内核以较低睿频运行。

全新英特尔® 至强® 金牌 6328HL、6328H 和 5320H 处理器支持英特尔® SST-CP 和英特尔® SST-TF 功能。

第三代英特尔®至强®可扩展处理器 SKU

如需了解最新信息,请访问 intel.cn/xeon 或 ark.intel.com

处理器标识	内核数/ 线程数	TDP (瓦)	处理器缓存	处理器基础频率 (GHz)	最大睿频速率 (GHz)	缓存 (MB)	内存容量 H/HL SKU 后缀
铂金 8380HL 处理器 铂金 8380H 处理器	28/56	250	38.5 MB	2.9	4.3	38.5	1.12 TB / 4.5 TB
铂金 8376HL 处理器 铂金 8376H 处理器	28/56	205	35.75 MB	2.6	4.3	38.5	1.12 TB / 4.5 TB
铂金 8354H 处理器	18/36	205	24.75 MB	3.1	4.3	24.75	1.12 TB
铂金 8353H 处理器	18/36	150	24.75 MB	2.5	3.8	24.75	1.12 TB
金牌 6348H 处理器	24/48	165	33 MB	2.3	4.2	33	1.12 TB
金牌 6328HL 处理器 金牌 6328H 处理器	16/32	165	22 MB	2.8	4.3	22	1.12 TB / 4.5 TB
金牌 5320H 处理器	20/40	150	27.5 MB	2.4	4.2	27.5	1.12 TB
金牌 5318H 处理器	18/36	150	24.75 MB	2.5	3.8	24.75	1.12 TB

	英特尔 [®] 至强 [®] 金牌处理器 (5300 系列)	英特尔 [®] 至强 [®] 金牌处理器 (6300 系列)	英特尔 [®] 至强 [®] 铂金处理器 (8300 系列)
支持的最大内核数	20 核	24核	28 核
支持的最高频率	3.3 GHz	3.7 GHz	4.3 GHz
支持的 CPU 路数	多达 4 个	多达 4 个	多达8个
英特尔® 超级通道互联(英特尔® UPI)	6	6	6
英特尔® UPI 速度	10.4 GT/s	10.4 GT/s	10.4 GT/s
英特尔 [®] 高级矢量扩展 512 技术(英特尔 [®] AVX-512 技术)	1 FMA	2 FMA	2 FMA
支持的最高内存速度 (DDR4)	2666 MT/s	2933 MT/s	3200 MT/s (1 DPC) 2933 MT/s (2 DPC)
每路支持的最高内存容量 [®]	1.12 TB	1.12 TB, 4.5 TB	1.12 TB, 4.5 TB
采用矢量神经网络指令 (VNNI) 和 16 位 Brain Floating Point (bfloat16) 数字格式处理功能的英特尔®深度学习加速 (英特尔®DL Boost)	•	•	•
英特尔® 傲腾™ 持久内存模组支持(仅适用于四路设计)	•	•	•
英特尔® Omni-Path 架构(独立式 PCI Express 卡)	•	•	•
英特尔® QuickAssist 技术(集成在某些芯片组中)	•	•	•
英特尔® QuickAssist 技术支持(独立式 PCI Express 卡)	•	•	•
英特尔® 傲腾™ 固态盘支持	•	•	•
英特尔® 数据中心级 SSD 支持 (3D NAND)			•
PCI Express 3(每个处理器支持 48 个通道)			•
英特尔® QuickData 技术 (CBDMA)			
非透明桥 (NTB)			•
英特尔® 睿频加速技术 2.0	•		
英特尔®超线程技术(英特尔®HT技术)			•
节点控制器支持			•
可靠性、可用性和可维护性 (RAS)	标准	高级	高级
英特尔®确保运行技术		•	•
英特尔 [®] Speed Select 技术 (英特尔 [®] SST) - 功能包括英特尔 [®] SST-Core Power (SST-CP) 和英特尔 [®] SST-Turbo Frequency (SST-TF)			
英特尔® Infrastructure Management 技术(英特尔® IMT)	•	•	•
英特尔®资源调配技术(英特尔® RDT)	•	•	•
英特尔® 卷管理设备(英特尔® VMD)	•	•	•
英特尔® 虚拟化技术(英特尔® VT)	•	•	•
英特尔® Speed Shift 技术	•	•	•
英特尔 [®] 节点管理器 4.0	•	•	•
基于模式的执行控制	•	•	•
时间戳计数器扩展 (TSC) 虚拟化	•	•	•
面向数据中心的英特尔® Security Libraries (英特尔® ISecL-DC)	•	•	•
英特尔 [®] 高级矢量扩展 512 技术(英特尔 [®] AVX-512 技术)	•	•	•
集成英特尔® QAT 的英特尔® 密钥保护技术 (KPT)	•	•	•
英特尔® Platform Trust 技术 (PTT)	•	•	•
可一键激活 (One-Touch Activation, OTA) 的英特尔 [®] 可信执行技术(英特尔 [®] TXT)	•	•	•
产品名称 SKU	压缩	加密	RSA
英特尔® C621A 芯片组 LBG-1G	N/A	N/A	N/A
英特尔 [®] C627A 芯片组 (搭载英特尔 [®] QuickAssist 技术)	~65 Gb/s	100 Gb/s	100K 次操作/秒
英特尔 [®] C629A 芯片组 (搭载英特尔 [®] QuickAssist 技术)	~75-80 GB/s	100 Gb/s	N/A

[°]仅在某些特定 SKU 型号中受到支持。

一般提示和法律声明

性能测试中使用的软件和工作负荷可能仅在英特尔微处理器上进行了性能优化。

诸如 SYSmark 和 MobileMark 等测试均系基于特定计算机系统、硬件、软件、操作系统及功能。上述任何要素的变动都有可能导致测试结果的变化。请参考其他信息及性能测试(包括结合其他产品使用时的运行性能)以对目标产品进行全面评估。要了解更全面的信息,请访问 www.intel.cn/benchmarks。

性能测试结果基于截至配置中所示日期的测试,且可能并未反映所有公开可用的安全更新。有关配置详细信息,请参阅备份。没有任何产品或组件是绝对安全的。 您的成本和结果可能会有所不同。

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。

1. 与上一代产品相比,集成了支持 BF16 的英特尔®深度学习加速技术的第三代英特尔®至强®可扩展处理器在 ResNet50 图像分类吞吐量方面,人工智能训练性能提升高达 1.93 倍——新配置: 单节点,搭载 4 颗第三代英特尔® 至强® 铂金 8380H 处理器(预生产 28 C,250 W)的英特尔®参考平台 (Cooper City),总内存 384 GB(24 插槽/16 GB/3200),ucode 0x700001b,启用 HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔® 固态盘 800 GB OS 驱动器,ResNet-50v1.5吞吐量,https://github.com/Intel-tensorflow/tensorflow-bbf16/base,commit#828738642760358b388dBf615ded0c213f10c99a,Modelzoo: https://github.com/IntelAl/models/-b v1.6.1,Imagenet 数据集,oneDNN 1.4,BF16,BS=512。英特尔于 2020 年 5 月 18 日测试。基准配置:单节点,搭载 4 颗英特尔® 至强® 铂金 8280 处理器的英特尔®参考平台(Lightning Ridge),总内存 768 GB(24 插槽/32 GB/2933),ucode 0x4002f00,启用 HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔® 固态盘 800 GB OS 驱动器,ResNet-50 v1.5 吞吐量,https://github.com/Intel-tensorflow/tensorflow-b bf16/base,commit#828738642760358b388dBf615ded0c213f10c99a,Modelzoo:https://github.com/IntelAl/models/-b v1.6.1,Imagenet 数据集,oneDNN 1.4,FP32,BS=512。英特尔于 2020 年 5 月 18 日测试。

2. 与5年前的四路平台相比,在使用全新第三代英特尔[®] 至强[®] 可扩展处理器时,云端数据分析应用模型的性能提升高达 1.92 倍 - 新配置: 单节点,搭载 4 颗第三代英特尔[®] 至强[®] 铂金 8380H 处理器(预生产 28 C,250 W)的英特尔[®] 参考平台 (Cooper City),总内存 1536 GB(48 插槽/32 GB/3200 @2933),微代码 0x700001b,启用 HT,启用睿频,并配有 Ubuntu 18.04.4 LTS, 5.3.0-53-generic,1 块英特尔[®] 固态盘 240 GB OS 驱动器,4 个 P4610 3.2 TB PCIe NVME, 4 个 40 GbE x710 双端口,CloudXPRT vCP——数据分析、Kubernetes、Docker、Kafka、MinIO、Prometheus 和 XGBoost 工作负载,Higgs 数据集,英特尔于 2020 年 5 月 27 日测试。基准配置:单节点,搭载 4 颗英特尔[®] E7-8890 v3 处理器的英特尔[®] 参考平台 (Brickland),总内存 1024 GB(64 插槽/16 GB/1600),微代码 0x0000016,启用 HT,启用睿频,并配有 Ubuntu 18.04.4 LTS, 5.3.0-53-generic,1 块英特尔[®] 固态盘 400 GB OS 驱动器,4 个 P3700 2 TB PCIe NVME, 4 个 40 GbE x710 双端口,CloudXPRT vCP——数据分析、Kubernetes、Docker、Kafka、MinIO、Prometheus 和 XGBoost 工作负载,Higgs 数据集,英特尔于 2020 年 5 月 27 日测试。

3. 没有任何产品或组件是绝对安全的。

4. 基准配置:单节点,搭载 1 颗英特尔® 至强® 8280L 28 C @ 2.7 GHz 处理器的 Neon City,单个持久性内存模组配置(6 个 32 GB DDR4 DRAM; 1 个 {128 GB, 256 GB, 512 GB} 英特尔® 傲腾™ 持久内存 100 系列模组,15 W), ucode 版本: 04002F00,用以运行 Fedora 29 内核 5.1.18-200.fc29.x86_64 和 App-Direct 模式下的 3.8 版本 MLC。数据来源:2020ww18_CPX_BPS_DI。英特尔于 2020 年 4 月 27 日测试。新配置:单节点,搭载 1 颗英特尔® 至强® 预生产 CPX6 28 C @ 2.9 GHz 处理器的 Cooper City,单个持久性内存模块配置(6 个 32 GB DDR4 DRAM; 1 个 {128 GB, 256 GB, 512 GB} 英特尔® 傲腾™ 持久内存 200 系列模块,15 W), ucode 预生产,用以运行 Fedora 29 内核 5.1.18-200.fc29.x86_64 和 App-Direct 模式下的 3.8 版本 MLC。数据来源:2020ww18_CPX_BPS_BG。英特尔于 2020 年 3 月 31 日测试。

5. 与上一代产品相比,集成了支持 BF16 的英特尔®深度学习加速技术的第三代英特尔®至强®可扩展处理器在 ResNet50 图像分类吞吐量方面,人工智能推理性能提升高达 1.87 倍——新配置: 单节点,搭载 4 颗第三代英特尔® 至强® 铂金 8380H 处理器 (预生产 28 C, 250 W) 的英特尔® 参考平台 (Cooper City) 总内存为 384 GB (24 插槽/16 GB/3200) ,ucode 0x700001b,启用 HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔® 固态盘 800 GB OS 驱动器,ResNet-50v1.5吞吐量,https://github.com/Intel-tensorflow/tensorflow-b bf16/base,commit#828738642760358b388d8f615ded0c213f10c99a,Modelzoo: https://github.com/IntelAl/models/ -b v1.6.1,Imagenet 数据集,oneDNN 1.4,BF16,BS=56,4 项实例,28 内核实例,英特尔于 2020 年 5 月 18 日测试。基准配置:单节点,搭载 4 颗英特尔® 至强®铂金 8280 处理器的英特尔® 参考平台 (Lightning Ridge),总内存 768 GB (24 插槽/32 GB/2933),ucode 0x4002f00,启用HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔® 固态盘 800 GB OS 驱动器,ResNet-50 v 1.5 吞吐量,https://github.com/Intel-tensorflow/tensorflow -b bf16/base,commit#828738642760358b388d8f615ded0c213f10c99a,Modelzoo: https://github.com/IntelAl/models/ -b v1.6.1,Imagenet 数据集,oneDNN 1.4,FP32,BS=56,4 项实例,28 内核/实例,英特尔于 2020 年 5 月 18 日测试。

6. 与上一代产品相比,集成了支持 BF16 的英特尔® 深度学习加速技术的第三代英特尔® 至强® 可扩展处理器在 BERT 自然语言处理吞吐量方面,人工智能训练性能提升高达 1.7 倍——新配置:单节点,搭载 4 颗第三代英特尔® 至强® 铂金 8380H 处理器(预生产 28C,250W)的英特尔® 参考平台 (Cooper City),总内存 384 GB (24 插槽/16 GB/3200),ucode 0x700001b,启用 HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔® 固态盘 800 GB OS 驱动器,BERT-Large (QA) 吞吐量,https://github.com/Intel-tensorflow/tensorflow-b bf16/base,commit#828738642760358b388d8f615ded0c213f10c99a,Model-zoo: https://github.com/IntelAl/models/-b v1.6.1,Squad 1.1 数据集,oneDNN 1.4,BF16,BS=12,英特尔于 2020 年 5 月 18 日测试。基准配置:单节点,搭载 4 颗英特尔® 至强® 铂金 8280 处理器的英特尔®参考平台 (Lightning Ridge),总内存 768 GB (24 插槽/32 GB/2933),ucode ucode 0x4002f00,启用 HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔® 固态盘 800 GB OS 驱动器,BERT-Large (QA) 吞吐量,https://github.com/Inteltensorflow/tensorflow-b bf16/base,commit#828738642760358b388d8f615ded0c213f10c99a,Modelzoo:https://github.com/IntelAl/models/-bv1.6.1,Squad 1.1 数据集,oneDNN 1.4,FP32,BS=12,英特尔于 2020 年 5 月 18 日测试。

7. 与上一代产品相比,集成了支持 BF16 的英特尔®深度学习加速技术的第三代英特尔®至强®可扩展处理器在 BERT 自然语言处理吞吐量方面,人工智能训练性能提升高达 1.9 倍——新配置:单节点,搭载 4 颗第三代英特尔® 至强® 铂金 8380H 处理器(预生产 28C,250W)的英特尔®参考平台 (Cooper City),总内存 384 GB (24 插槽/16 GB/3200),ucode 0x700001b,启用 HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔® 固态盘 800 GB OS 驱动器,BERT-Large (QA) 吞吐量,https://github.com/Intel-tensorflow/tensorflow -b bf16/base,commit#828738642760358b388d8f615de d0c213f10c99a,Modelzoo: https://github.com/IntelAl/models/ -b v1.6.1,Squad 1.1 数据集,oneDNN 1.4,BF16,BS=32,4 项实例,28 内核/实例,英特尔于 2020 年 5 月 18 日测试。基准配置:单节点,搭载 4 颗英特尔® 至强®铂金 8280 处理器的英特尔®参考平台 (Lightning Ridge),总内存 768 GB(24 插槽/32 GB/2933),ucode 0x4002f00,启用 HT,启用睿频,并配有 Ubuntu 20.04 LTS,Linux 5.4.0-26,28,29-generic,英特尔®固态盘 800 GB OS 驱动器,BERT-Large (QA) 吞吐量,https://github.com/Intel-tensorflow/tensorflow -b bf16/base,commit#828738642760358b388d8f615ded0c213f10c99a,Modelzoo: https://github.com/IntelAl/models/-b v1.6.1,Squad 1.1 数据集,oneDNN 1.4,FP32,BS=32,4 项实例,28 内核/实例,英特尔于 2020 年 5 月 18 日测试。

8. 与 5 年前的四路平台相比,在使用全新第三代英特尔°至强°可扩展平台时,每分钟处理的 OLTP 数据库事务量多出达 1.98 倍——新配置: 单节点,搭载 4 颗第三代 英特尔°至强°铂金 8380H 处理器(预生产 28 C,250 W)的英特尔°参考平台 (Cooper City),总内存 768 GB(24 插槽/32 GB/3200),微代码 0x700001b,启用 HT,启用睿频,并配有 Redhat 8.1,4.18.0-147.3.1.el8_1.x86_64,1 个英特尔° 固态盘 240 GB OS 驱动器,2 个用于 DATA 的 6.4T P4610,2 个用于 REDO 的 3.2T P4610,1Gbps NIC,HammerDB 3.2,常用商业数据库,英特尔于 2020 年 5 月 13 日测试。基准配置: 单节点,搭载 4 颗英特尔°至强°处理器 E7-8890 v3 的英特尔°参考平台 (Brickland),总内存 1024 GB(64 插槽/16 GB/1600),微代码 0x16,启用 HT,启用睿频,并配有 Redhat 8.1,4.18.0-147.3.1.el8_1.x86_64,1 个英特尔° 固态盘 800 GB OS 驱动器,1 个 用于 DATA 的 1.6T P3700,1 个 用于 REDO 的 1.6T P3700,1Gbps NIC,HammerDB 3.2,常用商业数据库,英特尔于 2020 年 4 月 20 日测试。

