

基于第四代英特尔® 至强® 可扩展平台的 浪潮信息大数据·智算一体机解决方案

浪潮信息

“为了帮助用户构建卓越的智算基础设施，浪潮信息与英特尔在硬件层面开展了广泛的合作，本次推动的新一代大数据·智算一体机融合了双方的技术优势，不仅能够加速大数据分析、AI模型训练与推理等负载，还可支持用户实现基础资源的统一管理，进一步提升资源利用率。我们将借助该解决方案，帮助更多客户构建高效、敏捷的大数据分析与人工智能平台，推动数字化转型进程。”

— 魏健

浪潮信息产品方案开发部

挑战

利用大数据技术与 AI 技术激活数据要素潜能成为产业数字化转型的重要发展方向。随着数字化转型战略的深入，企业一方面需要构建卓越的大数据与 AI 平台以支撑业务创新，另一方面也需要化解以下挑战：

- **如何进一步提升大数据处理与 AI 性能：** 企业的数据规模正在快速扩张，AI 模型日趋复杂，在数据处理、模型训练和模型推理等方面带来了巨大的算力需求。
- **如何实现基础资源的统一管理与数据的充分共享：** 大数据与 AI 平台分散部署的模式通常会导致两个平台的基础资源难以实现统一管理，使得资源无法得到充分利用和便捷地实现数据共享。
- **如何实现便捷化产品实施：** 传统 IT 基础设施的实施涉及硬件准备、网络配置、资源规划、软件部署验证等问题，通常需要较长的交付时间。
- **如何控制基础设施的 TCO：** 大数据和人工智能系统的建设涉及软硬件的系统设计、集成、安装、部署、调试与运维，以及大数据和 AI 软件系统的调试优化等工作，涉及巨大的资金和人力成本。
- **如何实现数据自动化存储管理：** 由于数据量快速增长，大数据平台面临着存储数据量巨大、数据增量较高、数据类型混杂等问题，这需要大数据平台提供强大的数据管理功能，对处于冷温热等不同生命周期的数据进行自动分层管理。

解决方案概述

当前，利用大数据技术与 AI 技术激活数据要素潜能成为产业数字化转型的重要方向。IDC 报告显示，全球到 2025 年数据量将达到 175 ZB¹，年复合增长率高达 26%²。如何提升大数据与 AI 系统的性能与效率，帮助客户高效处理巨量增长的数据，满足客户在 IT 基础设施方面对计算性能、可靠性、可扩展性、兼容性等的需求，同时赋能信息和应用系统智慧，成为 IT 基础设施厂商面临的重要挑战。

¹ IDC. "The Digitization of the World From Edge to Core." Sponsored by Seagate. November 2018. [seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf](https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf)

² IBM. "3D NAND Technology – Implications to Enterprise Storage Applications." 2015 Flash Memory Summit. 2014. flashmemorysummit.com/English/Collaterals/Proceedings/2015/20150811_FM12_Yoon.pdf

在此背景下，构建大数据与人工智能（AI）融合的大数据智算一体化平台成为组织的迫切需求。一体化平台能够在强化面向海量数据的处理能力的同时，赋能深度学习、自然语言处理等创新技术的应用，以获得更加高效、精确的数据洞察，推动数字化转型战略的成功实施。在此基础上，组织将能够解决算力优化、软硬件融合、资源管理与运维等挑战，有效推动数字化技术创新。

为帮助组织更有效地利用大数据与 AI 进行数字化转型，浪潮信息推出了基于英特尔® 架构的大数据·智算一体机解决方案。该解决方案以浪潮信息 SRDC (Smart Rack Data Center) 一体化平台为基础，采用了第四代英特尔® 至强® 可扩展处理器、英特尔® 以太网适配器等高性能硬件，并集成部署浪潮信息大数据平台 Insight、英特尔® NativeSQL、英特尔® BigDL (Analytics Zoo)、智能存储管理 (Smart Storage Management, SSM) 等软件系统，旨在通过软硬件一体化集成，形成软硬件紧密耦合的资源池，构建计算、存储、网络和 AI 应用紧密融合的高性能大数据平台。第四代英特尔® 至强® 可扩展处理器内置了英特尔® 高级矩阵扩展 (英特尔® AMX) 加速器，可以显著提升 AI 推理性能。

浪潮信息大数据·智算一体机解决方案帮助用户重构了融合大数据和 AI 系统的 IT 系统集成部署架构，提供了一套高性能、高可靠、高扩展、易部署的大数据 AI 融合一体化平台，降低了 IT 基础设施的总体拥有成本 (TCO)。

浪潮信息大数据·智算一体机解决方案

浪潮信息新一代大数据·智算一体机解决方案基于浪潮信息大数据平台 Insight，融合大数据与人工智能组件，与浪潮信息 SRDC 一体化平台实现工厂化安装部署，旨在构建人工智能深度学习端到端管理体系，为用户建立大数据分析与人机智能模型训练推理的深度融合平台，实现智算一体化部署交付，降低用户大数据和人工智能平台的 TCO。

浪潮信息新一代大数据·智算一体机支持第四代英特尔® 至强® 可扩展处理器，不仅能够大幅提升大数据业务场景处理性能，还能够通过内置的英特尔® AMX 为 AI 工作负载中的训练和推理提供显著的性能提升。该平台采用预集成、预测试、预优化等措施，显著降低了系统交付和运维复杂度，提升了系统稳定性，实现了卓越性能和一站式交付部署能力。此外，它还能够针对不同的场景进行定制配置，为客户提供性价比最优的解决方案。

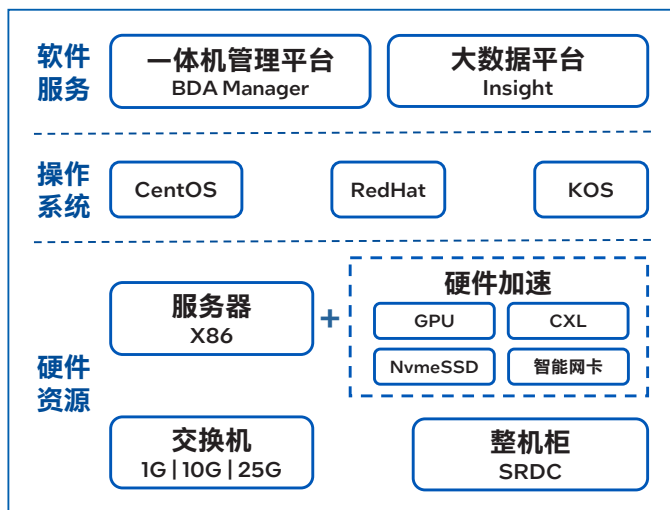


图 1. 浪潮信息新一代大数据·智算一体机解决方案架构

- 第四代英特尔® 至强® 可扩展处理器：**通过创新架构增加了每个时钟周期的指令，每个插槽多达 60 个核心，支持 8 通道 DDR5 内存，有效提升了内存带宽与速度，并通过 PCIe 5.0 (80 个通道) 实现了更高的 PCIe 带宽提升。第四代英特尔® 至强® 可扩展处理器提供了现代性能和安全性，可根据用户的业务需求进行扩展。借助内置的加速器，用户可以在 AI、分析、云和微服务、网络、数据库、存储等工作负载中获得优化的性能。相比第三代处理器，第四代英特尔® 至强® 可扩展处理器实现了 1.53 倍³ 的代际性能提升。通过与强大的生态系统相结合，第四代英特尔® 至强® 可扩展处理器能够帮助用户构建更加高效、安全的基础设施。
- 英特尔® 以太网适配器 E810-DA2：**具备 100/25GbE 性能，支持单个或双端口连接，在 PCIe 4.0 x16 插槽中提供了出色的性能，并支持应用程序设备队列 (ADQ)、动态设备个性化 (DDP)、RDMA iWARP 和 RoCEv2 等各种高级功能，能够有效满足各种工作负载对于带宽和时延的严苛要求。

浪潮信息新一代大数据·智算一体机搭载了浪潮信息云海 Insight 大数据平台，该平台具备数据智能化分析挖掘、多数据源协同计算、大规模集群一键式部署、支持隐私计算、异构平台混合部署与资源精细化管理等领先技术特性。平台提供“一键式”安装部署功能，让大数据·智算一体机达到“开箱即用”的程度，帮助用户加快应用系统上线时间，同时降低系统管理员日常运维管理复杂度。

³ 详细配置信息请访问：[intel.com/processorclaims](https://www.intel.com/processorclaims)，选择“第四代英特尔® 至强® 可扩展处理器”，查看编号“G1”。实际性能受使用情况、配置和其他因素的差异影响。

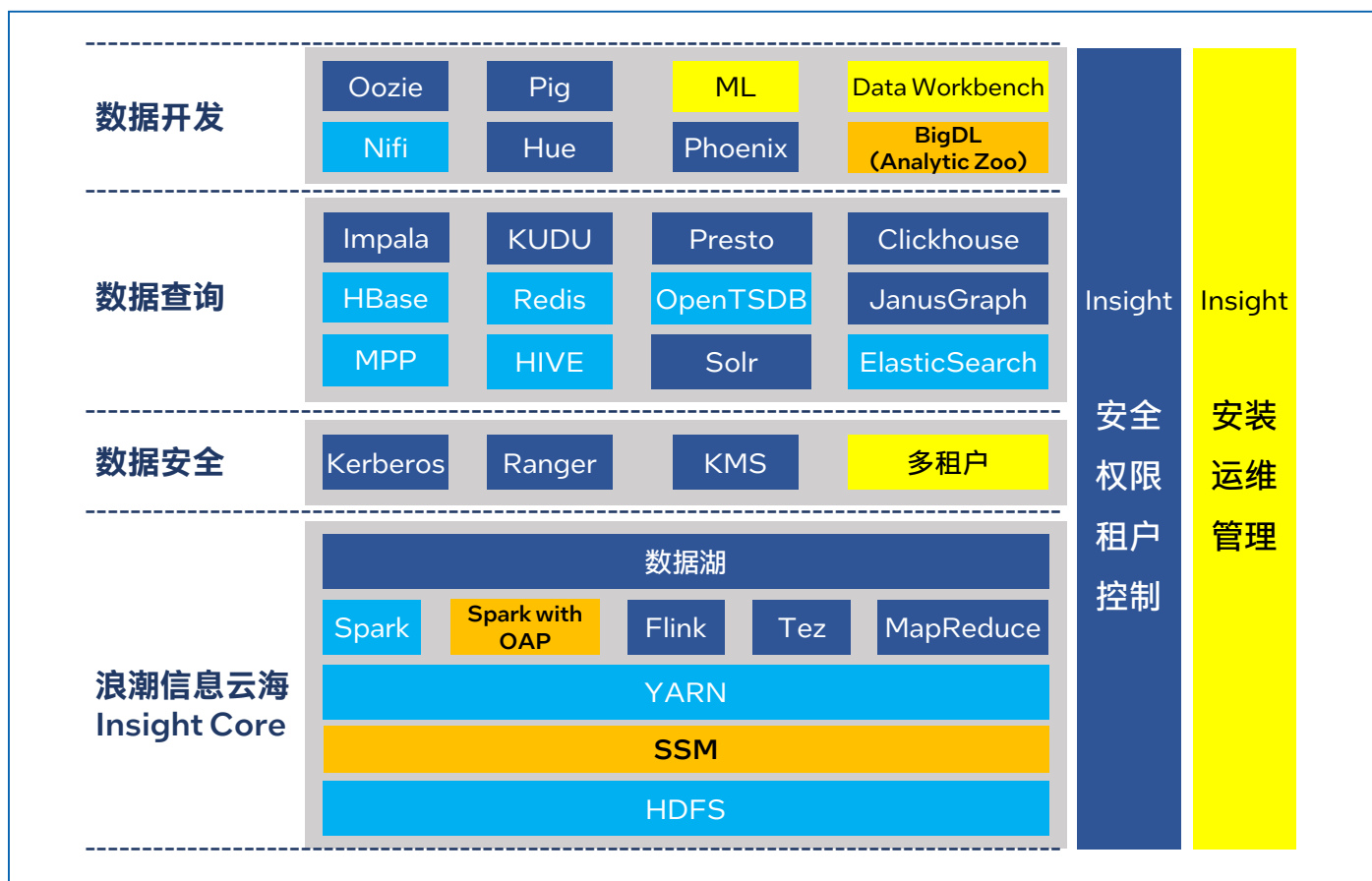


图 2. 浪潮信息云海 Insight 大数据平台技术架构

浪潮信息与英特尔面向大数据与 AI 实现深度优化

在浪潮信息云海 Insight 大数据平台，以及英特尔新一代软硬件产品与技术的基础上，浪潮信息与英特尔围绕 AI 工作负载加速、智能存储管理、端到端智算统一管道的构建、大规模数据分析引擎性能优化等方面，进行了深度的合作。

使用英特尔® AMX 加速 AI 工作负载

英特尔® AMX 是内置于第四代英特尔® 至强® 可扩展处理器中的加速器，可优化深度学习 (DL) 训练和推理工作负载。借助英特尔® AMX，第四代英特尔® 至强® 可扩展处理器可在优化通用计算和 AI 工作负载间快速转换。开发人员可以编写非 AI 功能代码来利用处理器的指令集架构 (ISA)，也可编写 AI 功能代码，以充分发挥英特尔® AMX 指令集的优势。

英特尔® AMX 架构如图 3 所示，主要由两部分组件构成：第一部分为 TILE，由 8 个 1KB 大小的 2D 寄存器组成，可存储大数据块；第二部分为平铺矩阵乘法 (TMUL)，它是与 TILE 连接的加速引擎，可执行用于 AI 的矩阵乘法计算。英特尔® AMX 支持两种数据类型：INT8 和 BF16，两者均可用于 AI 工作负载所需的矩阵乘法运算。

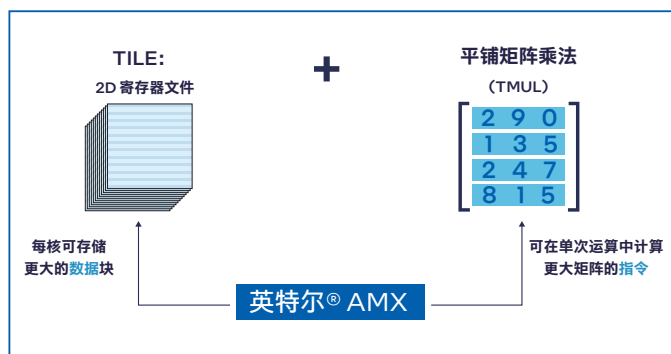


图 3. 英特尔® AMX 架构

为了验证英特尔® AMX 带来的性能提升，浪潮信息在基于第四代英特尔® 至强® 可扩展处理器的浪潮信息新一代大数据·智算一体机上，采用 ResNet50 模型进行了 AI 推理性能测试。

测试数据如图 4 所示，与 FP32 数据类型相比，在通过英特尔® AMX 将数据类型转化为 BF16 与 INT8 之后，模型推理性能分别提升到之前的 5.2 倍与 9.26 倍⁴。

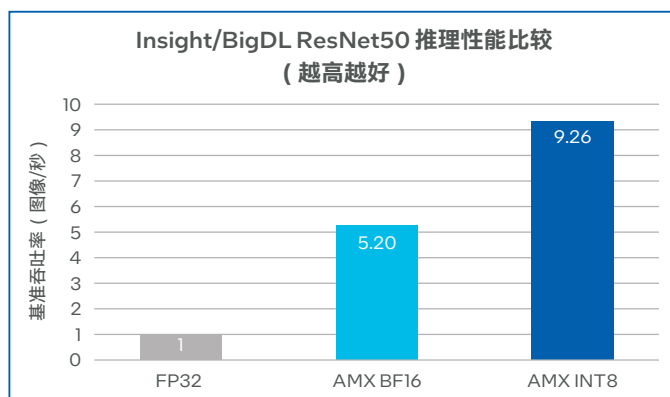


图 4. ResNet50 模型推理性能比较⁵

英特尔 NativeSQL 携手浪潮信息 Insight 助力 Spark 焕“芯”算力

浪潮信息新一代大数据·智算一体机搭载了英特尔® AVX512 指令集加持的现代化矢量引擎 NativeSQL，最高可提升 Spark 任务性能达 7 倍⁶。NativeSQL 是云海 Insight 大数据平台为 SQL 分析场景引入的计算内核硬科技，作为 Spark 插件式 SQL 计算引擎，它基于 C++ 语言重构 Spark SQL 执行层，开箱即用，与英特尔技术相结合，能够提供更强算力，提升用户在复杂业务处

理，如标签画像、进销存分析、推荐系统、数据 ETL 等大批量数据的离线计算场景的效率。

在 TPC-H 测试中，Spark NativeSQL 总体性能相比开源版本 Spark3.1 提升 75%，单查询最高提速 193%，在 TPC-DS 基准中，单查询性能最高超过 700%，且在多轮测试中随着基准数据量的增大，性能提升效果呈现上升趋势⁷。

通过英特尔® SSM 释放智能存储管理潜力

浪潮信息新一代大数据·智算一体机搭载了英特尔® 智能存储管理组件 (英特尔® SSM)，为用户提供存储优化和数据管理方案。SSM 是英特尔在 Apache Hadoop 生态系统中发起和维护的存储管理方案，提供了对异构存储硬件的支持，可利用英特尔® 傲腾™ 固态硬盘实现动态优化数据存储，实现远端存储与 HDFS 本地备份统一管理。

英特尔® SSM 与浪潮信息云海 Insight 大数据平台的无缝结合实现了面向分布式存储的智能管理，无需外界触发就可以智能进行数据管理，进一步细化管理粒度，而且仅需一次规则定制，即可完成全数据生命周期管理。

集成 BigDL (Analytics Zoo) 构建端到端智算安全可信流水线

为了简化端到端的大数据分析和人工智能融合开发和部署，浪潮信息新一代大数据·智算一体机在 Insight 中集成了英特尔开源的人工智能解决方案平台 BigDL (Analytics Zoo)，可以将大数据预处理、模型训练、模型推理与现有的大数据处理 workflows 整合在一起。

此外，浪潮信息新一代大数据·智算一体机还在 Insight 中集成了基于英特尔® SGX/TDX 技术的英特尔® BigDL PPML (Privacy Preserving Machine Learning, 隐私保护的机器学习) 框架，为端到端大数据人工智能流水线提供了一个受信任的集群环境。英特尔® SGX/TDX 能够帮助用户构建基于硬件的数据中心可信

^{4,5}截至 2022 年 9 月浪潮信息联合英特尔开展的测试。测试配置：双路英特尔® 至强® 铂金 8455C 处理器 @ 2.0 GHz，48 核，启用超线程，启用睿频加速技术，1024 GB 总内存 (16x64 GB 4800 MT/s)，1.7 TB 固态硬盘，Ubuntu 22.04.1 LTS。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.com/PerformanceIndex

^{6,7}数据援引自浪潮信息内部测试结果。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。

执行环境 (TEE)。英特尔® SGX 通过将特权代码排除在受信任的范围之外，具有较小的攻击面，能够更有效地抵御多种类型的攻击；而英特尔® TDX 通过提供硬件隔离的安全虚拟机，可以有效避免来自非信任域的攻击，同时更好地兼容现有的应用和软件生态。通过采用英特尔® SGX/TDX 技术，并结合多种安全技术，英特尔® BigDL-PPML 允许用户以安全可信的方式运行未经修改的大数据分析、机器学习和 AI 等应用，并且提供了隐私集合求交、安全聚合和联邦学习等隐私计算方案。

浪潮信息新一代大数据·智算一体机能够在浪潮信息一体化机柜平台 SRDC 得到最优配置与交付。SRDC 面向智慧数据中心而设计，为用户提供了集成服务器、交换机、PDU 等 IT 设备，以及配电单元、UPS、电池包、动环监控和基础资源监控 (ISPIM) 等配套硬件设备和软件的一体化机柜平台。SRDC 以整机柜为最小产品颗粒度，在工厂集成业务所需机柜内设备，根据用户需求集成预装应用软件，可实现 IT 设备快速部署及业务快速上线。

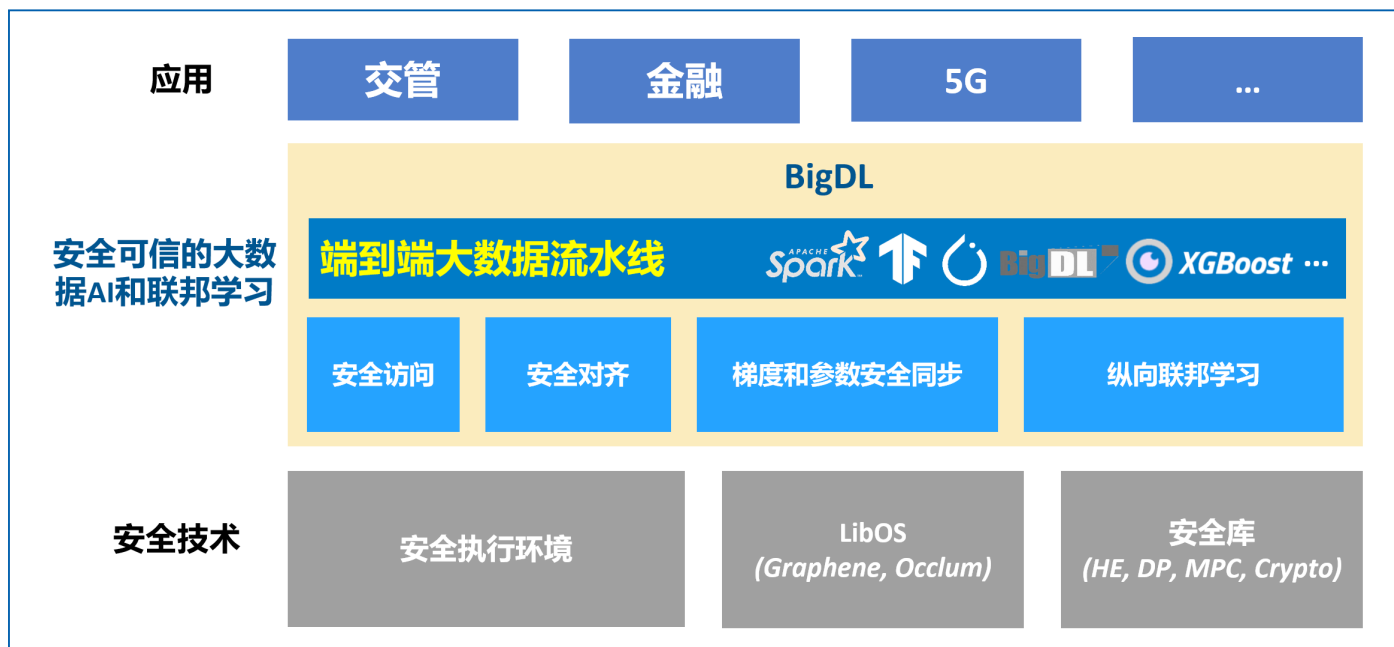


图 5. 安全可信的端到端的大数据分析和人工智能融合架构

收益：帮助用户构建大数据与 AI 统一平台

通过部署浪潮信息新一代大数据·智算一体机，用户能够构建大数据与 AI 统一平台，有效解决大数据与 AI 平台的资源与数据无法统一管理的问题，提升计算、存储、网络等资源利用率，同时节约在软硬件配置、性能优化等方面的时间与成本。

卓越性能加速大数据与 AI 负载运行

浪潮信息新一代大数据·智算一体机实现了软硬件深度整合，可以充分发挥第四代英特尔® 至强® 可扩展处理器的优势，全面提高端到端计算效率。

提升 IT 基础实施交付效率

浪潮信息新一代大数据·智算一体机通过软硬件深度适配调优、预安装、预测试、预集成等措施，实现开箱即用，支持快速交付部署，简化交付流程。

实现软硬件全生命周期管理

浪潮信息新一代大数据·智算一体机构建了软硬一体化监控平台，可以实现一站式、可视化运维管理，纳管集群软硬件资源，实现对硬件、软件全生命周期管理。

保证数据的安全和可靠

浪潮信息新一代大数据·智算一体机具有出色的可靠性与可用性。全冗余硬件架构全面消除单点故障，同时成熟可靠的多副本技术实现了数据级别的冗余，保证数据的安全与可靠。

展望：AI 与大数据深度融合助力数字化转型

基于英特尔® 架构的浪潮信息大数据·智算一体机解决方案融合了浪潮信息智算基础设施、浪潮信息云海 Insight 大数据平台，并充分利用了内置英特尔® AMX 的第四代英特尔® 至强® 可扩展处理器等硬件，以及英特尔相关软件技术，为客户提供了基于英特尔® 架构的全栈式大数据与 AI 解决方案，不仅能够加快大数据分析 AI 的部署、降低开发门槛，还有利于提升模型训练、模型推理速度，降低 TCO。

未来，双方计划继续合作挖掘第四代英特尔® 至强® 可扩展处理器的潜力，例如利用英特尔® In-Memory Analytics Accelerator (英特尔® IAA) 优化内存占用和查询吞吐量，采用英特尔® Data Streaming Accelerator (英特尔® DSA) 跨多个核心高效进行

负载均衡，利用英特尔® QuickAssist 技术 (英特尔® QAT) 提升平台加密、压缩等性能。通过上述合作，双方将有望进一步提升浪潮信息大数据·智算一体机在 AI、数据分析、隐私计算、数据加解密等场景下的处理能力。

此外，双方还将面向未来智算基础设施的异构参考架构，进一步开展合作，通过英特尔® 至强® 可扩展处理器、NNP 芯片、FPGA、网络、内存和存储等基础硬件，以及各种数据库、人工智能平台、数据湖与开发工具，构建优化型的大数据与 AI 端到端安全可信解决方案，帮助用户更快速处理大规模的数据，加速大数据、AI 的开发与应用。

关于浪潮信息

浪潮信息是全球领先的 IT 基础设施产品、方案和服务提供商，拥有 8 个研发中心、10 个生产基地、26 个分支机构，业务遍及 120 多个国家和地区。秉承“计算力就是生产力，智算力就是创新力”的理念，浪潮信息致力于推动智慧计算技术创新和应用，加速数实相融，让人们生活更美好，企业经营更高效，社会治理更完善，人与自然更和谐。浪潮信息坚持“伙伴第一”的原则，携手伙伴发展共融共生的计算产业生态，为客户提供全栈式、一体化的智算解决方案，加速行业智能化转型。

关于英特尔

英特尔 (NASDAQ: INTC) 作为行业引领者，创造改变世界的技术，推动全球进步并让生活丰富多彩。在摩尔定律的启迪下，我们不断致力于推进半导体设计与制造，帮助我们的客户应对最重大的挑战。通过将智能融入云、网络、边缘和各种计算设备，我们释放数据潜能，助力商业和社会变得更美好。如需了解英特尔创新的更多信息，请访问英特尔中国新闻中心 newsroom.intel.cn 以及官方网站 intel.cn。



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.com/PerformanceIndex

性能测试结果基于配置信息中显示的日期进行测试，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。

英特尔未做出任何明示和默示的保证，包括但不限于，关于适销性、适合特定目的及不侵权的默示保证，以及在履约过程、交易过程或贸易惯例中引起的任何保证。

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在美国和/或其他国家的商标。