

# 阻碍微服务的五大挑战

## 微服务的巨大影响及面临的诸多挑战

在云原生领域，微服务在开发人员中的普及率持续攀升，这并不奇怪。英特尔最近开展的一项研究显示，83% 的全新云原生应用以及软件即服务 (SaaS) 解决方案都在使用微服务<sup>1</sup>。

开发人员、基础设施运营商以及企业/机构的 IT 部门在采用基于微服务的应用时面临着重大挑战。企业和机构需要提高吞吐量，既符合时延服务级别协议 (SLA) 要求，又能够应对服务质量、基础设施开销和可观测性等微服务方面的关键挑战。

本白皮书概要介绍了微服务面临的五大关键挑战，并就英特尔为应对这些挑战所做的努力（包括您可以在自身所处的环境中利用的工具、功能和资源）进行了深度分析。

### 挑战一：性能和长尾时延



云原生、基于微服务的应用在提供可预测的响应时间以及满足严格的 SLA 要求方面压力巨大。减小时延（抖动）窗口和保证时延一致是成功的关键。从丢包到主机故障，各种因素均有可能导致异常值的产生。

#### 解决方案：英特尔的上游优化和深度软件分析

为了帮助企业 and 机构解决长尾时延和性能上的挑战，英特尔正在努力优化处理器，更好地满足基于微服务应用的新领域的需求。与此同时，英特尔还致力于通过深入分析软件工作负载（包括行业基准测试和代理、代表性工作负载和客户/合作伙伴工作负载），找出微服务架构内的瓶颈问题。基于所获得的分析结果，英特尔通过基础设施架构、整体软件改进和独特的英特尔® 加速器推动软件优化，帮助缓解性能和长尾时延问题。

英特尔软件工程师已构建了数种运行时与语言优化方案，为提升微服务工作负载的速度和效率助一臂之力。这些优化被推向上游并加进最新版本的通用开源语言库，让您能够更轻松地满足自身云原生应用的性能要求。

第四代英特尔® 至强® 可扩展处理器内置众多加速器，可为 AI、数据分析、网络、存储和科学计算等快速增长的工作负载提供性能和能效优势。在全新英特尔® 高级矩阵扩展（Intel® Advanced Matrix Extensions，英特尔® AMX）这一加速器的支持下，第四代英特尔® 至强® 可扩展处理器展现出更为出色的 AI 训练和推理性能。而其他无缝内置于处理器中的加速器则可分别加速数据传输和压缩，从而实现更高的网络性能；可提升查询吞吐量，从而实现响应更加迅捷的数据分析；同时还可将调度和队列管理任务卸载到云平台的多个内核上，实现工作负载动态平衡。为实现新的内置加速器功能，英特尔为技术生态系统提供了操作系统级软件、库和 API 支持。

#### 精确时间管理 (PTM)

##### 提升跨多个计算节点进行超大规模部署的云服务和应用的性能

Google、Meta、Amazon 和 Microsoft 等重要客户都认识到了大幅提升跨数据中心的时间同步准确性的价值，其中很多客户已经部署了依赖精确网络时间同步（如 IEEE 1588-2019 标准等）的分布式应用。

第四代英特尔® 至强® 可扩展处理器采用精确时间管理 (PTM) 机制，它在服务器内将精确的网络时间从 PCIe 网络设备一直扩展到 CPU 上运行的应用软件，从而完成服务器之间可实现的数量级性能提升。

[了解更多有关采用时间敏感网络实现自动化系统的信息 >](#)

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<sup>2</sup>。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® 运行时和语言优化包括：

**Java 与 JVM：**提高科学仿真、金融分析、AI/深度学习、3D 建模和分析、图像和音频/视频处理、密码操作、数据压缩及其他密集型工作负载的性能。英特尔® 高级矢量扩展 512 ( Intel® Advanced Vector Extensions 512, 英特尔® AVX-512 ) 是最新的 x86 矢量指令集，包含多达两个融合乘加 (FMA) 单元和其他优化，可为要求严苛的计算任务提升性能。

**Go：**密码操作优化、压缩 ( Gzip、Brotli )、HTML 转义。

**PHP：**面向英特尔® 架构 ( iTLB、dTLB、NUMA ) 的整体性能优化，支持控制流强制技术以提升安全性。

**Node.js：**实现 JSON 解析矢量化，以减少微服务通信开销。

## 挑战二：分布式通信



微服务的实现取决于任意数量节点、应用和服务间的高效通信。如果计算平台没有足够高的性能与安全性，就可能会出现影响用户满意度和数据隐私的重大问题。

### 解决方案：使用全新第四代英特尔® 至强® 可扩展处理器的内置加速器来提升并帮助保护微服务通信

利用英特尔® 数据流加速器 ( Intel® Data Streaming Accelerator, 英特尔® DSA )，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® DSA 旨在卸载数据中心级部署中会产生开销的常见数据传输任务，有助于加速数据在 CPU、内存和缓存以及各种附加内存、存储和网络设备中的传输。

在多核英特尔® 至强® 可扩展处理器上启用英特尔® 动态负载均衡器 ( Intel® Dynamic Load Balancer, 英特尔® DLB ) 可提升与网络数据处理相关的系统性能。英特尔® DLB 能够在多个 CPU 内核/线程之间有效地分配网络处理负载，并能随着系统负载的变化，将网络数据动态地分配到多个 CPU 内核上进行处理。此外，英特尔® DLB 还可以复位同时在 CPU 内核上处理的多个网络数据包的顺序。

### 服务网络优化

英特尔® 数据保护与压缩加速技术 ( Intel® QuickAssist Technology, 英特尔® QAT ) 通过加速密码操作、密钥保护和数据压缩，降低系统资源消耗。该内置加速器可以卸载加密和解密任务，从而帮助释放处理器内核资源，助力系统服务更多客户端。

#### 网络层性能提升

英特尔可为众多常见的云原生网络构件提供优化：

#### 网络性能：

ADQ、Calico、Cilium、eBPF、Ethernet Operator、Kubernetes 中的 AF\_XDP、Kubernetes 中的 IPU/IPDK

#### RPC 性能：

gRPC、Thrift、Kraton

#### 总体网络强化：

硬件卸载、内核网络 ( eBPF、AF\_XDP )

### 挑战三：编排和基础设施开销



由于涉及大量节点且所在环境瞬息万变，因此微服务在管理上面临着诸多挑战。要想解决这些问题，为这些环境和应用提供技术支持的团队需要能从计算平台中获得合适的功能。

#### 解决方案：利用英特尔内置或集成的发现、自动化及优化功能来简化管理

为尽可能减少对核心引擎的修改，Kubernetes 提供了支持扩展功能的插件架构和机制。英特尔新增一些扩展功能，可让应用更好地利用加速器、功耗管理和调度/工作负载分配插件。此外，英特尔还提供了可自动配置其中多种扩展功能的 Operator。这样一来，开发人员便可以更轻松地使用这些功能。

英特尔® QAT 和英特尔® 软件防护扩展 ( Intel® Software Guard Extensions, 英特尔® SGX ) 已启用，可以通过 Kubernetes 轻松使用。另外，英特尔® 架构旨在提供可以帮助您优化工作负载调度的相关功能和 NUMA 拓扑以及平台遥测技术，后者可以提供有助于提高利用率和降低总体拥有成本 (TCO) 所需的可执行洞察。

例如，拓扑感知调度可确保将出于时延考虑而必须避免跨处理器或处理器内核边界的进程锁定在正确的内核中。对 CPU 负载和温度等遥测方面的深度分析结果有助于实现自动化、高效工作负载分配和能耗优化。

#### 为什么使用面向 Kubernetes 的英特尔® 技术？

##### 集成功能

- 利用“节点功能配置文件发现”
- 发现节点功能并暴露给控制面板以优化分配
- 限制预定义配置文件数量以简化功能的使用

##### Operator 与自动化

- 通过自动启用 ( 设备插件、功耗管理、网络 ) 简化功能部署

##### 工作负载调度

- 将 NUMA 拓扑感知扩展到调度程序中
- 利用平台遥测技术所获洞察提高利用率和优化 TCO

### 挑战四：易用性和开发人员工作效率



微服务开发带来了独特挑战。因此，英特尔提供很多专门的资源，帮助您构建微服务应用。英特尔致力于同开发者社区协作，以简化流程、加速运营并优化结果。

#### 解决方案：使用专门的英特尔® 开发人员资源简化开发工作

##### 利用我们的硬件与软件构件提升您的工作效率

英特尔提供全面的软件与硬件构件产品组合和生态系统，从编排和高性能网络到处理器、以太网控制器等，面面俱到。

##### 利用容器化环境参考架构提升性能

英特尔为裸机和虚拟容器化环境提供系统级蓝图，助力开发流程的简化和性能优化，并为 5G 和 Open RAN 部署提供多种经验证的堆栈供选择。

##### 使用英特尔的体验套件，挖掘英特尔多种功能的潜能

集成的英特尔® 体验套件提供了开发人员充分利用英特尔® 功能与加速技术所需的各种资源，其中包括指南、分步视频和训练模块，帮助您全面利用我们的技术。

参见 [网络和云边缘参考系统架构产品组合用户手册](#)，获得有关使用第三代和第四代英特尔® 至强® 可扩展处理器及英特尔® 至强® D 处理器平台在特定网络位置部署经验证的系统的指导。

## 挑战五：可观测性



对于基于微服务且具有高度分布及不断移动特点的应用而言，保持对环境中所发生情况的完全可视和感知至关重要。深度的可观测性和可视性有助于您更轻松地对需调整和优化，以满足严苛的 SLA 要求。

### 解决方案：基于英特尔® 架构进一步挖掘深度可视化功能

#### 充分利用英特尔® 性能计数器

启用裸机上的英特尔® 性能计数器 (Intel® Performance Counters)，并选择公有云环境，可以轻松了解微服务工作负载的性能表现。例如，您可以用这些计数器提供的信息检查各个工作负载的长尾时延，并找出导致其未能满足 SLA 要求的常见因素。这种做法可以轻松实现历史数据的可视化并对其进行分析，帮助您找到问题并予以解决。

#### 借助遥测收集器 (Telemetry Collection Agent) 获得可执行洞察

英特尔® 服务器遥测技术覆盖的领域广泛，包括利用率、功耗、故障检测和性能。为了帮助用户从这类信息中获取有意义的洞察，英特尔创建了一套遥测技术报告，提供有关服务器当前状态的可执行数据。将上述洞察与有关软件和服务运行情况的数据结合在一起，就能更全面地了解网络功能。

---

## 结论：英特尔致力于让您的微服务大获成功

当您寻求开辟新领域并推出基于微服务的出色应用时，英特尔将利用自身技术和生态系统优势帮助您取得成功。我们的上游优化、集成功能和更优性能可助您应对当下关键的微服务挑战。

## 想要了解更多有关基于英特尔® 架构的微服务的信息？

[不断了解微服务及英特尔如何帮助增强微服务的信息](#) >

[查看边缘视频分析微服务的 Docker 映像文件](#) >

[观看 Tech Field Day Showcase 中有关基于英特尔® 技术的云实例的微服务优化视频](#) >

[快速了解零信任 5G 网络安全微服务](#) >



#### 一般提示和法律声明

1. 英特尔微服务深度分析研究，2021年6月。
2. 配置详情请见以下网址的 [I26-130]: <https://edc.intel.com/content/www/cn/zh/products/performance/benchmarks/3rd-generation-intel-xeon-scalable-processors/>。

英特尔技术可能需要启用硬件、软件或激活服务。具体成本和结果可能不同。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。

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](#)。

© 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。

1022/MH/CMD/PDF