White Paper

Microservices

Five Challenges That Hinder Microservices

The enormous impact and sizable challenges of microservices

In a cloud-native world, it's no surprise that microservices adoption continues to rise among developers. A recent study conducted by Intel reveals that 83 percent of all new cloud-native applications and software-as-a-service (SaaS) solutions are using microservices.[1]

Developers, infrastructure operators, and organizational IT departments face significant challenges as they embrace microservices-based applications. Organizations need optimized throughput that stays with latency service level agreements (SLAs) while addressing key microservices challenges such as quality of service, infrastructure overhead, and observability.

This white paper outlines five of the most critical microservices challenges and provides insights on how Intel is working to solve them—including the tools, features, and resources you can take advantage of in your environment.

Challenge 1: Performance and tail latencies

Cloud-native, microservices-based applications are under pressure to deliver predictable response times and meet tight SLAs. Achieving a consistent, small window of possible latency times is essential to success. Outliers can be caused by any number of factors, from packet loss to host failure.

Solution: Intel's upstream optimizations and deep software analysis

To help solve the tail latency and performance challenge, Intel is working to optimize processors for the new world of microservices-based applications. Intel is also working to identify bottleneck points within microservices architectures by deeply analyzing software workloads—including industry benchmarks and proxy, representative, and customer/partner workloads. These insights fuel software optimizations through infrastructure set architecture, general software improvements, and unique Intel® accelerators that help mitigate performance and tail latency issues.

Intel software engineers have built several runtime and language optimizations to help improve speed and efficiency for microservices workloads. These are pushed upstream for acceptance into the latest versions of common open source language repositories—making it easier for you to meet performance expectations for your cloud-native applications.

4th Gen Intel® Xeon® Scalable processors have the most built-in accelerators of any CPU on the market to deliver performance and power-efficiency advantages across the fastest-growing workload types in AI, analytics, networking, storage, and HPC. With all-new accelerated matrix multiply operations, 4th Gen Intel Xeon Scalable processors have exceptional AI training and inference performance. Other seamlessly integrated accelerators speed up data movement and compression for faster networking, boost query throughput for more responsive analytics, and offload scheduling and queue management to

Precision time measurement (PTM)

Increase the performance of cloud services and applications hyperscaled across many compute nodes

Leading customers, including Google, Meta, Amazon, and Microsoft, recognize the value of dramatically improving time synchronization accuracy across data centers, and many have deployed distributed applications that rely on precision network time synchronization (IEEE Std. 1588-2019 and similar).

4th Gen Intel® Xeon® Scalable processors include Precision Time Measurement (PTM) that extends precision network time within the server from PCIe network device all the way up to the application software running on the CPU, complementing the orders-of-magnitude improvement achievable between servers.

Learn more about adopting Time-Sensitive Networking for automation systems ›

dynamically balance loads across multiple cores. To enable new built-in accelerator features, Intel supports the technology ecosystem with OS-level software, libraries, and APIs.

With built-in accelerators and software optimizations, previous-generation Intel Xeon Scalable processors have been shown to deliver leading performance per watt on targeted real-world workloads.[2] This results in more efficient CPU utilization, lower electricity consumption, and higher ROI while helping businesses achieve their sustainability goals.

Intel® runtime and language optimizations include:

Java and JVM: Accelerate performance for scientific simulations, financial analytics, AI/deep learning, 3D modeling and analysis, image and audio/video processing, cryptography, data compression, and other intensive workloads. Intel® Advanced Vector Extensions 512 (Intel® AVX-512) is the latest x86 vector instruction set, with up to two fused-multiply add (FMA) units and other optimizations to accelerate performance for your most demanding computational tasks.

Go: Cryptography optimization, compression (Gzip, Brtoli), HTML escaping.

PHP: General performance optimizations for Intel® architecture (iTLB, dTLB, NUMA), enabling control-flow enforcement technology for better security.

Node.js: JSON parsing vectorization to reduce microservices communication overhead.

Challenge 2: Distributed communications

Microservices depend on efficient communication between any number of nodes, applications, and services. Without adequate performance and security from your compute platform, you can face significant issues that impact user satisfaction and data privacy.

Solution: Use new 4th Gen Intel Xeon Scalable processor accelerators to enhance and help protect microservices communications

Drive high performance for storage, networking, and data-intensive workloads by improving streaming data movement and transformation operations with the Intel® Data Streaming Accelerator (Intel® DSA). It is designed to offload the most common data movement tasks that cause overhead in data center–scale deployments. Intel DSA helps speed up data movement across the CPU, memory, and caches, as well as all attached memory, storage, and network devices.

Improve the system performance related to handling network data on multicore Intel Xeon Scalable processors with the Intel® Dynamic Load Balancer (Intel® DLB). It enables the efficient distribution of network processing across multiple CPU cores/threads and dynamically distributes network data across multiple CPU cores for processing as the system load varies. Intel DLB also restores the order of networking data packets processed simultaneously on CPU cores.

Service mesh optimizations

Help reduce system resource consumption by providing accelerated cryptography, key protection, and data compression with Intel® QuickAssist Technology (Intel® QAT). By offloading encryption and decryption, this built-in accelerator helps free up processor cores and helps systems serve a larger number of clients.

Networking-layer performance improvements

Intel delivers optimizations for many of the most common cloud-native networking building blocks:

Network performance:
ADQ, Calico, Cilium, eBPF, Ethernet Operator, AF_XDP in Kubernetes, IPU/IPDK in Kubernetes

RPC performance:
gRPC, Thrift, Katran

General networking enhancements:
Hardware offload, kernel networking (eBPF, AF_XDP)

Challenge 3: Orchestration and infrastructure overhead

With massive numbers of nodes and highly dynamic environments, microservices present sizable management challenges. To solve them, the teams that support these environments and applications need the right features from their compute platform.

Solution: Simplify management with integrated Intel features for discovery, automation, and optimization

Kubernetes seeks to minimize changes to the core engine and therefore provides a plugin architecture and mechanisms to enable extensions. Intel adds extensions that allow applications to take advantage of accelerators, power management, and the scheduling/workload placement add-ons. Intel also delivers operators that automate the configuration of many of these extensions, making it easier for developers to use them.

Intel QAT and Intel® Software Guard Extensions (Intel® SGX) are enabled for easy consumption via Kubernetes. Plus, Intel® architecture is built to provide the capabilities and NUMA topology that can help you optimize workload scheduling—with platform telemetry that provides the actionable insights needed to improve utilization and enhance TCO.

For example, topology-aware scheduling ensures that processes that need to avoid crossing processor or processor-core boundaries for latency reasons are pinned to the right core. Telemetry insights such as CPU load and temperature allow for automation, workload placement efficiency, and energy optimization.

Why Intel for Kubernetes?

Integrated features

- Leverage node feature profile discovery
- Discover features on nodes and expose to control plane for optimized placement
- Simplify feature adoption with limited number of predefined profiles

Operators and automation

- Simplify deployment of features by automating enablement (device plugins, power management, networking)

Workload scheduling

- Expand awareness of NUMA topology into the scheduler
- Improve utilization and TCO by tapping into platform telemetry insights

Challenge 4: Ease of use and developer productivity

Microservices development presents unique challenges. That's why Intel provides purpose-built resources to help you build microservices applications. Intel is committed to working with the developer community to streamline processes, expedite operations, and optimize results.

Solution: Simplify development with purpose-built Intel® developer resources

Accelerate your efforts with our hardware and software building blocks

Intel offers a full portfolio and ecosystem of software and hardware building blocks—from orchestration and high-performance networking to processors, Ethernet controllers, and more.

Accelerate performance with reference architectures for containerized environments

By providing system-level blueprints for both bare-metal and virtualized containerized environments, Intel helps simplify development processes and optimize performance. Intel also offers validated stacks for 5G and Open RAN deployments, with multiple options available.

Get the most out of Intel capabilities with our experience kits

Integrated Intel® experience kits provide everything developers need to take full advantage of integrated Intel® features and accelerations. They include guides, step-by-step videos, and training modules that help you fully leverage our technologies.

See the **Network and Cloud Edge Reference System Architectures Portfolio User Manual** that has guided deployments of validated systems at specific network locations using 3rd and 4th Gen Intel® Xeon® Scalable processor and Intel® Xeon® D processor platforms.

Challenge 5: Observability

With highly distributed, constantly moving microservices-based applications, maintaining complete visibility and awareness of what's happening in your environment is critical. Deep observability and visibility make it easier to adjust and optimize as needed to meet demanding SLAs.

Solution: Unlock deeper visibility on Intel architecture

Take advantage of Intel performance counters

Activate Intel performance counters—available on bare-metal and select public cloud environments—to easily understand performance of microservices workloads. For example, you can use the information provided by these counters to examine tail latencies across your workloads and spot common factors that are causing it to violate an SLA. Historic data can easily be visualized and analyzed to help identify and resolve issues.

Gain actionable insights with Telemetry Collection Agent

Intel® server telemetry spans a vast number of domains including utilization, power consumption, fault detection, and performance. To offer meaningful insights from this information, Intel has created a portfolio of telemetry reports that provides actionable data about the current status of the server. Combining these insights with performance data about how the software and services are operating allows for a more holistic view of the network function.

Conclusion: Intel is dedicated to your microservices success

As you seek to break new ground and roll out amazing microservices-based applications, Intel is working across our technologies and ecosystem to enable your success. Our upstream optimizations, integrated features, and accelerated performance deliver what you need to solve today's most pressing microservices challenges.

Want to learn more about microservices on Intel architecture?

Continue learning about microservices and how Intel helps enhance them ›

Check out our docker image for an edge video analytics microservice ›

Watch the Tech Field Day Showcase on microservices optimization for Intel-based cloud instances ›

Get up to speed on security microservices for zero trust 5G networks ›

intel®

Notices and disclaimers

1. Intel Microservices Insights Study, June 2021.
2. 3rd Gen Intel Xeon Scalable Processor vs. AMD EPYC. See configuration details [126-130] at www.intel.com/3gen-xeon-config.

Intel® technologies may require enabled hardware, software, or service activation. Your costs and results may vary.
Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.
Availability of accelerators varies depending on SKU. Visit the Intel Product Specifications page for additional product details.

1022/MH/CMD/PDF