Develop and deploy high-performance cloud workloads faster

DevOps teams creating an e-commerce site typically start by downloading open source code from GitHub. For example, they might download open source code for NGINX, a popular web server for high-traffic websites. From there, the team can optimize the software for performance and package it for deployment.

But there is an alternate approach for developers using cloud instances with Intel hardware. This approach can significantly accelerate time to value: starting with Intel Optimized Cloud Stacks.

To create Intel Optimized Cloud Stacks, Intel starts with the same open source software available to developers on GitHub. Intel then creates application-specific software optimizations. These optimizations accelerate performance and are tuned for Intel hardware on specific cloud instances.

The optimizations take advantage of Intel’s advanced acceleration capabilities for cryptography, artificial intelligence (AI), and other CPU-intensive workloads. Intel then works with partners like Bitnami to build, compile, and package the software. Using Intel Optimized Cloud Stacks as a starting point for workloads like NGINX, WordPress, and MySQL can save developers a significant amount of time. It can also greatly enhance performance.

Figure 1 illustrates Intel Optimized Cloud Stacks for MySQL, an open source relational database management system (RDBMS). Intel starts with the standard open source image and enhances it with software optimizations, including security. The software is then tuned for an Intel instance on a target public cloud. Users choose either a single-node or a multi-node version, depending on their use case: static or dynamic web hosting, e-commerce, social media, banking applications, or others.

This paper offers software architects and developers information on the advantages of using Intel Optimized Cloud Stacks:

- How Intel Optimized Cloud Stacks differ from other packaged open source software in the cloud marketplaces
- Details on software configurations and optimizations for Intel components such as asynchronous OpenSSL and the Intel QuickAssist Technology (Intel QAT) engine
- Performance benefits of Intel Optimized Cloud Stacks with real-world data
- Tips for benchmarking performance

DevOps teams that start with Intel Optimized Cloud Stacks can save significant time and enhance performance.

How are Intel Optimized Cloud Stacks different?

Intel Optimized Cloud Stacks differ from other open source packages in several ways. To start, Intel uses unique software components, and it tunes these components to use the advanced instruction-set capabilities of Intel platforms. Intel also optimizes the components to use the capabilities of public cloud instances. By upgrading and tuning components, Intel enables organizations to deploy applications with higher performance and in less time than if they used the standard packages available on GitHub.
Tables 1 and 2 display optimized components included in Intel stacks. Table 1 outlines the Intel Optimized Cloud Stacks for WordPress, whereas Table 2 outlines the Intel Optimized Cloud Stacks for NGINX.

To enable higher-speed response times for websites, Intel uses asynchronous NGINX instead of standard NGINX in both the WordPress and NGINX images. Similarly, to help ensure communications security over networks while delivering high performance, Intel uses OpenSSL in asynchronous mode instead of standard OpenSSL.

Both the NGINX and WordPress offerings include Intel Multi-Buffer Crypto, which accelerates the packet processing of IP security (IPsec). IPsec is a framework of related protocols that help secure communications at the packet processing layer. The stacks also include the Intel QAT engine, which accelerates and compresses cryptographic workloads by offloading the data to hardware that optimizes those functions. Both of these components can help improve the performance of the NGINX engine, enabling data center servers to respond efficiently to high traffic levels.

**Table 1. Intel Optimized Cloud Stacks for WordPress**

<table>
<thead>
<tr>
<th>Component</th>
<th>Bitnami image optimized for 3rd Generation Intel® Xeon® Scalable processors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating system (OS) distribution</td>
<td>Debian 10 (dependent on the chosen virtual machine [VM])</td>
</tr>
<tr>
<td>Siege</td>
<td>v2.78</td>
</tr>
<tr>
<td>MySQL server</td>
<td>MySQL v15.1, distr 10.3.29-MariaDB</td>
</tr>
<tr>
<td>WordPress version</td>
<td>5.8</td>
</tr>
<tr>
<td>Zend Engine</td>
<td>Zend Engine v3.4.0 with Zend OPcache v7.4.19</td>
</tr>
<tr>
<td>PHP version</td>
<td>PHP 7.4.19 (command-line interface [CLI]) (built May 7, 2021, 11:03:55) (NTS)</td>
</tr>
<tr>
<td>Base NGINX version</td>
<td>NGINX/1.18.0</td>
</tr>
<tr>
<td>OpenSSL version</td>
<td>1.1.1d</td>
</tr>
<tr>
<td>Intel QAT engine</td>
<td>v0.6.5</td>
</tr>
<tr>
<td>Asynchronous NGINX</td>
<td>v0.4.5</td>
</tr>
<tr>
<td>Intel Multi-Buffer Crypto for IPsec</td>
<td>ippcp_2021.2</td>
</tr>
<tr>
<td>Library (Cryptomb)</td>
<td></td>
</tr>
<tr>
<td>IPsec MB</td>
<td>v1.0</td>
</tr>
</tbody>
</table>
Table 2. Intel Optimized Cloud Stacks for NGINX

<table>
<thead>
<tr>
<th>Component</th>
<th>Bitnami image optimized for 3rd Generation Intel Xeon Scalable processors</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS distribution</td>
<td>Debian 10 (dependent on the chosen VM)</td>
</tr>
<tr>
<td>Base NGINX version</td>
<td>NGINX/1.18.0</td>
</tr>
<tr>
<td>OpenSSL version</td>
<td>1.1.1d (included in the OS)</td>
</tr>
<tr>
<td>Intel QAT engine</td>
<td>v0.6.5</td>
</tr>
<tr>
<td>Asynchronous NGINX</td>
<td>v0.4.5</td>
</tr>
<tr>
<td>Intel Multi-Buffer Crypto for IPSec Library</td>
<td>ippcp_2021.2</td>
</tr>
<tr>
<td>IPsec MB</td>
<td>v1.0</td>
</tr>
</tbody>
</table>

Intel differentiates its Intel Optimized Cloud Stacks in the way it curates the components. They are selected to harness Intel's advanced hardware-acceleration capabilities for specific cloud instances.

Intel also manages the heavy lifting of building, compiling, and packaging the application software. This allows DevOps teams to easily launch Intel Optimized Cloud Stacks from the public CSP site. This is in contrast to running through multiple memory and CPU cycles preparing software packages.

Optimized software components

The following sections take a deeper look at some of the optimized software included in Intel Optimized Cloud Stacks and how these components can help increase performance.

Asynchronous OpenSSL

The OpenSSL project provides an open source implementation of the Secure Sockets Layer (SSL)/Transport Layer Security 3 (TLS3) protocols needed to authenticate and encrypt network connections between computers. SSL/TLS3 encrypt personal information sent by applications like e-mail and social media. Performance is critical because of ever-increasing amounts of traffic within data centers.

Asynchronous OpenSSL is an important contribution to this project that addresses this need for improved performance.

Asynchronous OpenSSL is a non-blocking approach that supports a parallel-processing model at the cryptographic level for SSL/TLS protocols. This, in turn, allows for other types of optimizations. Two major benefits of asynchronous OpenSSL are increased single-flow throughput, leading to high performance, and fewer contexts, thus reducing context-management overhead. By including asynchronous OpenSSL in Intel Optimized Cloud Stacks, developer applications can benefit from superior server performance.
Intel QuickAssist Technology (Intel QAT) engine for OpenSSL

OpenSSL performance can be further improved with the Intel QAT engine, which is supported by select Intel architecture platforms. Intel QAT has been expanded to provide software-based acceleration of cryptographic operations through instructions in the Intel Advanced Vector Extensions 512 (Intel AVX-512) family.² Again, by including Intel QAT in Intel Optimized Cloud Stacks, applications can benefit from enhanced server performance. The Intel QAT engine for OpenSSL is deployable today for on-premises users, and it can provide benefits in the cloud as CSPs roll out instances that expose it.

Asynchronous mode for NGINX

Open source NGINX is a lightweight, high-performance server that can serve content to a high volume of connections. It is used by high-demand websites because of its asynchronous architecture and its ability to handle heavy loads.

Intel extends NGINX by enabling it to work with asynchronous-mode OpenSSL. Intel also incorporates Intel QAT acceleration. Running NGINX with asynchronous OpenSSL contributes to better server performance.

Intel Multi-Buffer Crypto for IPsec Library

Intel Optimized Cloud Stacks use Intel Multi-Buffer Crypto for IPsec Library instructions that provide crypto acceleration of packet-processing frameworks like IPsec and TLS. Intel also provides algorithmic and software enhancements that enable operation chaining, advanced cryptographic pipelining, multi-buffering, and function-stitching innovations.³ By including Intel Multi-Buffer Crypto for IPsec Library in Intel Optimized Cloud Stacks, improved performance is possible.

Web server performance

You can use the TLS 1.2 version web server benchmark to compare crypto performance with and without the Intel QAT engine using IFMA crypto instructions. Intel testing demonstrates the advantage of using the Intel QAT engine. As shown in Figure 3, crypto acceleration was 7x faster for the NGINX TLS web server when the Intel QAT engine was used.⁴

![NGINX TLS 1.2 web server performance](chart)

**NGINX TLS 1.2 web server performance**

Normalized connections per second

(Higher is better)

<table>
<thead>
<tr>
<th></th>
<th>With crypto acceleration</th>
<th>Without crypto acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vCPU</td>
<td>6.3</td>
<td>1</td>
</tr>
<tr>
<td>16 vCPU</td>
<td>7.1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Figure 2.** TLS 1.2 web server crypto-acceleration results

AI analytics

Intel creates application-specific software optimizations for Intel hardware in use cases beyond web services. For example, the Intel oneAPI AI Analytics Toolkit (AI Kit) uses optimizations to accelerate end-to-end AI analytics workflows in the Python ecosystem.

As an example, Intel used two optimized software components from the AI Kit and trained a model to predict income based on education.⁵ The Intel Distribution of Modin was employed for data ingestion and extract, transform, and load (ETL) functions, while the Intel Extension for Scikit-learn was used for model training and prediction.

In Figure 3, you can see the optimized software components boost ETL by 38x, in addition to machine-learning (ML) prediction and fitting with ridge regression being improved 21x compared to using the stock libraries.⁶
Tips for benchmarking performance

When evaluating performance, be aware that benchmark tests can be deceiving. Some companies might claim that their processors run specific workloads in public clouds faster than on Intel processors. But these claims are often cherry-picked, focusing on workloads rarely encountered in a production environment. These “synthetic” benchmarks aren’t meaningful in the real world. These companies might also compare fully optimized Arm binaries against unoptimized x86 images. Consider requesting real-world performance results rather than relying on synthetic benchmarks.

A top-down microarchitecture similarity analysis delivers testing and validation that more closely resembles the real world. In Figure 4, production workloads cluster at the lower left, reflecting both front-end and back-end bound execution that results in higher sensitivity to both microarchitecture and memory latency.

Figure 3. Census dataset benchmark results

Figure 4. A top-down microarchitecture similarity analysis delivers testing and validation that more closely resembles the real world, as shown by the example use cases (light blue), application workloads (dark blue), and SPEC components (red)
The workload similarity analysis of CPU performance-monitor data shown in Figure 4 also highlights the substantial differences in performance when comparing various real-world workload proxies to benchmark tests mathematically. Standard Performance Evaluation Corporation (SPEC) benchmarks cluster components into microbenchmarks that lack the front-end boundedness of real applications. Even non-SPEC benchmarks must be tweaked so that they can reflect a proxy of the real workload. SPEC benchmarks and similar microbenchmarks primarily use CPU-bound indicators that are gated by clock speed and/or core count. As a result, SPEC benchmarks tend to measure only the frequency product from the simple cores. But there are other factors—cache and memory latencies, non-uniform memory access (NUMA), and input/output (I/O), for example—that are far more meaningful in real-world usages. As Figure 4 illustrates, a proxy-based approach is the better way to gauge the benefits of Intel Optimized Cloud Stacks.

Gain higher performance in less time

Intel Optimized Cloud Stacks are designed to help developers and architects get the most from the cloud. They can save development and deployment time. Their built-in software optimizations and tuning enable workloads to harness cutting-edge Intel technologies and public cloud capabilities. As a result, organizations can meet performance demands successfully, whatever they might be.

Learn more about getting the most from your cloud

Learn more about Intel tools for cloud optimization and management:
intel.com/content/www/us/en/cloud-computing/cloud-tools.html

Find optimized cloud stacks for NGINX and WordPress on public cloud marketplaces:

NGINX Open Source for Intel packaged by Bitnami:

AWS Marketplace:
https://aws.amazon.com/marketplace/pp/prodview-2ie7l4qcbccj

Azure Marketplace:

GCP Marketplace:

WordPress for Intel packaged by Bitnami:

AWS Marketplace:

Azure Marketplace:

GCP Marketplace:
Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.
Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.
Your costs and results may vary.
Intel technologies may require enabled hardware, software or service activation.
© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.