



Accelerating Cloud Applications with Intel® Xeon® Scalable Processors and Intel® FPGAs

Cloud Service Providers (CSPs) accelerate data-intensive applications and workloads for high throughput, excellent power efficiency, and fast response times using Intel® Xeon® Scalable processors and Intel® FPGAs

Industry Strategic Challenges

The number of connected devices and the amount of data that they produce continues to increase. According to Gartner, there will be 20.4 billion connected things worldwide by 2020¹. The rate at which data arrives from these devices into data centers also continues to grow.

At the same time, Cloud Service Providers (CSPs) face constraints in their data center space, power, and cooling and must work within tight margins. Any solutions CSPs deploy must be power efficient, and support low latency so they can keep up with the surge in big data. Additionally, they must be programmable and versatile: fixed silicon accelerators are unable to match the variety of workloads that CSPs face.

By alternating Intel® FPGAs, a wide range of workloads can be accelerated to accommodate this increased data, and new demands for analyzing it, within the constraints of the CSP's data center.

Accelerating Applications with FPGAs

FPGAs are silicon devices that can be dynamically reprogrammed with a data path that exactly matches your workloads, such as [data analytics](#), image inference, encryption and compression. This versatility enables you to provision fast processing power efficiency, and low-latency service – potentially lowering your total cost of ownership, and maximizing compute capacity within the power, space and cooling constraints of your data centers.

Traditionally, FPGAs require deep domain expertise to program, but the Intel® Acceleration Stack for Intel® Xeon® CPU with FPGAs simplifies the development flow and enables rapid deployment across the data center. Intel is collaborating (partner is a forbidden word at Intel) with FPGA developers, server OEMs, virtualization platform providers, operating system vendors and system integrators to enable customers to efficiently develop and operationalize their infrastructure.

Solution Benefits

- **Ease of deployment.** The Intel® Programmable Acceleration Card (Intel® PAC) provides an Intel® FPGA in a PCIe-based card. Validated servers incorporating the card are available from several leading OEMs.
- **Standardization.** The Intel® Acceleration Stack for Intel® Xeon® CPU with FPGAs defines standardized interfaces that FPGA developers and development and operations teams can use to hot-swap accelerators and enable application portability.
- **Programmability.** The Acceleration Stack abstracts away much of the complexity of programming FPGAs, making it easier for Tier 1 Cloud Service Providers (CSPs) and large cloud-based businesses to develop their own accelerators.
- **Off-the-shelf availability.** Accelize's AccelStore* offers a library of predefined accelerators that next-wave CSPs can reuse, with charge-back mechanisms built-in, making it easier to create an on-demand acceleration service. These accelerators are developed by an ecosystem of collaborators with deep workload processing expertise and validated for use by Accelize.

The FPGA can be directly transacted on with thread-safe APIs from within virtual machines and containers, or indirectly through standard libraries that have been extended to alternative FPGAs. This relieves developers of the need to craft customized drivers and debugging interfaces, enabling them to focus on their core expertise – algorithm development – and develop their solutions faster, and with greater confidence.

Modern data centers can consist of thousands of servers with virtual machines or containers running on them, all of which are managed as a large resource pool by an orchestrator. Intel's Acceleration Stack for FPGAs enables orchestrators such as Kubernetes* and OpenStack* to discover and manage FPGAs like any other compute resource, and hot swap accelerator cores into them in seconds. This capability is available for both service provisioning in a public cloud, and application cluster managers in private clouds. Since the versatility of Intel FPGAs allows for almost any workload to be accelerated, almost any server can be run more efficiently, and the performance improvement could enable new service offerings and what matters most – a better customer experience.

To further simplify the use of FPGAs in servers, Intel has developed an Intel® Programmable Acceleration Card with Intel® Arria® 10 GX FPGA (Intel® PAC with Intel® Arria® 10 GX FPGA). The card contains a networking interface for accelerating workloads such as streaming analytics and video transcoding, together with dedicated banks of DDR4 memory with error correction. The Intel PAC with Intel Arria 10 GX FPGA also includes the FPGA Interface Manager, and seamlessly pairs with an Intel Xeon processor over the PCIe bus. Intel is collaborating with hardware accelerator developers, server OEMs, systems integrators and independent software vendors (ISVs) to enhance compatibility and the out-of-box experience.

Solution Summary

The Intel® Acceleration Stack for FPGAs is a new architecture that allows software developers to use the power of FPGAs much more easily than before. A core component of the Acceleration Stack is the FPGA Interface Manager, which provides performance-optimized connectivity between an Intel FPGA and an Intel® Xeon® processor.

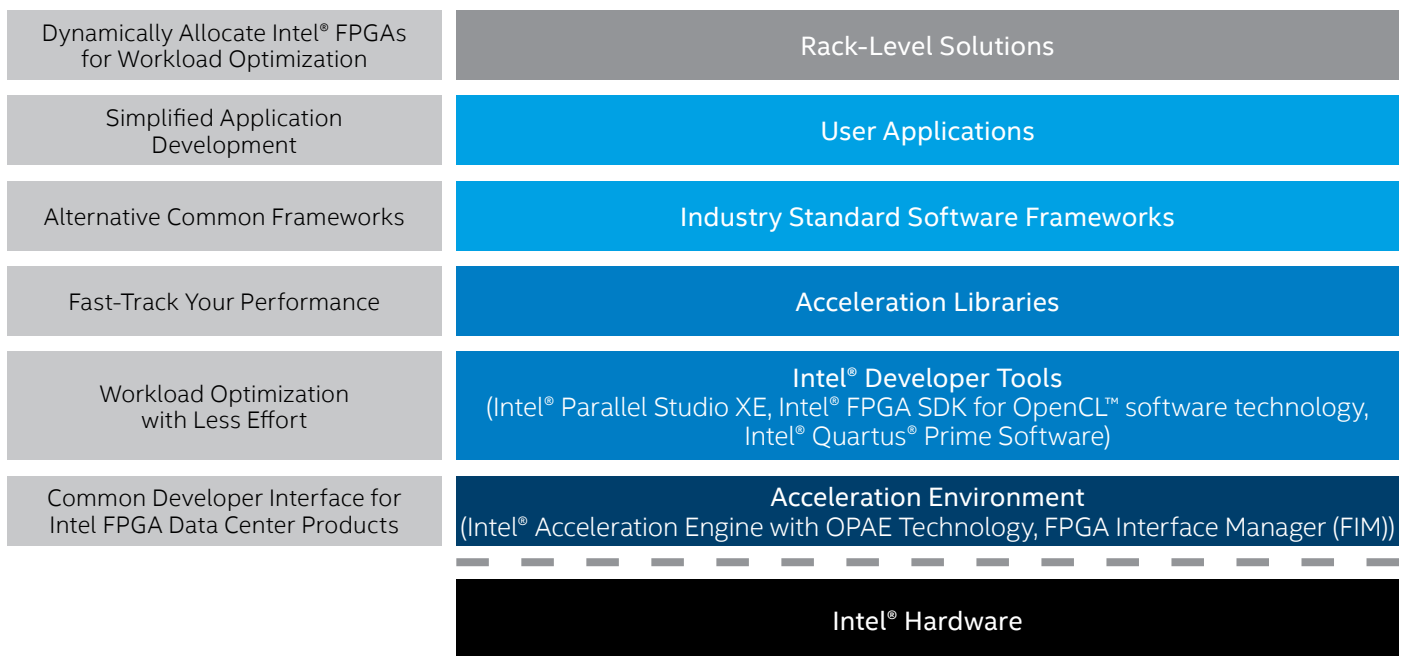


Figure 1. Acceleration Stack for Intel® Xeon® CPU with FPGAs

Customer Benefits for CSPs

CSPs can help accelerate their customers' workloads by using FPGAs in their data centers. Most of the Tier 1 CSPs have already made announcements about using FPGAs in their data centers. For example, Alibaba Cloud provides cloud-based access to Intel FPGAs in a pay-as-you-go model, using the Acceleration Stack for Intel Xeon CPU with FPGAs. Users will have access to off-the-shelf accelerator configurations for genomics, machine learning, data analytics, cyber security, financial computation and video transcoding. "Intel FPGAs offer us a more cost-effective way to accelerate cloud-based application performance for our customers who are running business applications and demanding data and scientific workloads," said Jin Li, vice president of Alibaba Cloud. "Another key value of FPGAs is that they provide high performance at low power, and the flexibility for managing diverse computing workloads."

Intel FPGAs, backed by the Acceleration Stack and Accelize AccelStore*, form the basis of OVH's Acceleration as a Service offering. Accelerating functions such as finding and replacing text saves users time and money, and frees up cycles on OVH's server for higher-value processes. OVH is also able to increase its operating margins by charging for the amount of data processed, rather than execution time.

Conclusion

The Intel Acceleration Stack for Intel Xeon CPU with FPGAs enables CSPs to more easily develop and deploy accelerated workloads in the data centers. FPGA platform development complexity is abstracted away, allowing CSPs to focus on the unique aspects of their applications and platforms, and deploy faster. Through Intel's collaborator ecosystem, CSPs can access off-the-shelf accelerators for use on the FPGAs, and deploy with robust metering and monitoring infrastructure.

Learn More

- [Intel® FPGAs Resource Center](#)
- [Intel® FPGAs for Data Analytics](#)
- [Intel® FPGA Acceleration Hub](#)
- [White paper: Simplify Software Integration for FPGA Accelerators with OPAAE](#)
- [Accelize website](#)
- [Intel® Xeon® Scalable platform](#)

Find the solution that is right for your organization. Contact your Intel representative or visit intel.com/cloud.



¹ <https://www.gartner.com/newsroom/id/3598917>

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Intel, the Intel logo, Xeon, Arria, OpenCL, and Quartus are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.