

CASE STUDY

High Performance Computing (HPC) in the Cloud
2nd Generation Intel® Xeon® Scalable Processors
Intel® Xeon® Platinum Processors



Astera Labs Novel Cloud-only Approach Leads to Faster, Better Chip Designs

Quick access to the latest Custom Intel® Xeon® Scalable processors in Amazon EC2 z1d instances proves invaluable

Technology Highlights:

- Amazon EC2 z1d instances—Custom Intel® Xeon® Scalable processor
- Amazon EC2 C5 instances—2nd Generation Intel® Xeon® Scalable processor
- Amazon EC2 R5 instances—Intel® Xeon® Platinum 8000 series processors



Executive Summary

[Astera Labs, Inc.](#), a fabless semiconductor company headquartered in California's Silicon Valley, develops purpose-built connectivity solutions for data-centric systems. The company's product portfolio includes system-aware semiconductor integrated circuits, boards, and services to enable robust PCIe connectivity. The company went from being a brand-new startup to having a fully designed next-generation PCIe 5.0 chipset in less than a year using a novel cloud-only business model running on Amazon Web Services (AWS). As an AWS user, Astera Labs was among the first semiconductor companies to take advantage of the custom Intel® Xeon® Scalable processor in the Amazon EC2 z1d instances. By relying on the cloud for all of its high performance computing (HPC) needs, Astera Labs estimates it reduced the development time for its initial semiconductor chip offering by 30–40% compared to using a traditional model while still ensuring a high-quality product that meets the stringent needs of their demanding customers.

Challenge

Astera Labs, with a passionate vision to be “the trusted partners to distribute data in intelligent systems,” started with a completely clean slate in late 2017. Founded by industry veterans, the company had a strong chip design background, but it lacked expertise in running IT systems and managing electronic design automation (EDA). After carefully evaluating the tradeoffs between on-premises, hybrid or cloud-based IT, Astera Labs chose a 100% public cloud approach based on options for outsourcing support and priorities around business flexibility and compute elasticity among other factors.

According to Jitendra Mohan, CEO of Astera Labs, although his team was eager to get moving with the cloud-based approach it didn't know how to get started. “Astera reached out to AWS, who in turn referred us to [Six Nines IT](#), a Premier AWS Consulting Partner,” said Mohan. “Six Nines explained our different options on AWS and based on their advice we did our own study and put together our preferred approach. Initially, we weren't making any effort to set up the most efficient file server possible and those sorts of things. We just wanted to be able to run simulations like we were used to and get moving on our designs. It only took a couple of weeks from having literally nothing other than our laptops to being able to run some simulations.”

Solution

Jason Cutrer, founder and president of Six Nines IT, said that Six Nines helped Astera Labs architect, configure and maintain their end-to-end chip design environment, which is 100% in AWS. “One of the advantages on partnering with Six Nines is that we have a lot of industry experience with EDA-specific and HPC workloads,

Case Study | Astera Labs Novel Cloud-only Approach Leads to Faster, Better Chip Designs

so we were able to greatly accelerate their configuration and time-to-market with an infrastructure to assist in their design and simulations,” explained Cutrer. “As we engaged, we made recommendations to use a variety of Intel-powered AWS instance types, which helped accelerate Astera’s timelines.”

As part of its initial efforts in the cloud, Astera Labs worked on designing a new semiconductor chip. “For simulations, we predominantly used Amazon EC2 C5 and R5 instances along with EC2 z1d instances once they became available. For the backend, we rely on different types of machines, including the z1d, R5 and R4 instances. We also use other instance types for general purpose tasks,” stated Mohan.

Depending on size, the C5 instances include either custom 2nd generation Intel Xeon Scalable processors, 2nd generation Intel Xeon Scalable Processors or the 1st generation Intel Xeon Platinum 8000 series processors. The instances with custom 2nd generation Intel Xeon Scalable processors have a sustained all-core turbo frequency of 3.6 GHz and maximum single core turbo frequency of 3.9 GHz, making them ideal workhorses for simulations. The others can achieve a sustained all core Turbo frequency of up to 3.4 GHz, and single core turbo up to 3.5 GHz using Intel® Turbo Boost Technology.

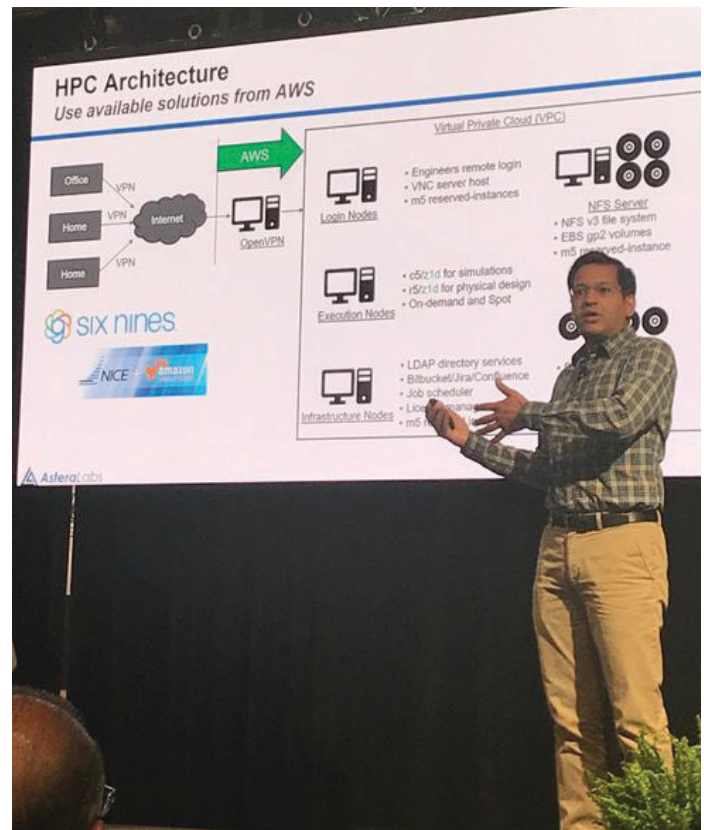
With custom Intel Xeon Scalable processors, the z1d instances are among the fastest in the cloud. They provide high single thread performance and sustained core frequency of up to 4.0 GHz. Even higher performance is possible with the help of Intel® Turbo Boost technology, which accelerates processor and graphics performance for peak loads when cores are operating below specific power, current and temperature thresholds. “When we put together our initial architecture, the z1d instances weren’t even released yet. Once they became available, we used them heavily for simulations and backend as well,” said Mohan.

Result

The z1d instances are specifically designed for the EDA industry, so being able to start using them as soon as they were available in the public cloud was a huge benefit. “As the saying goes, ‘time is money’ and we started using the z1d because of processing speed. The availability of z1d instances and the opportunity to run many jobs in parallel gave us the ability to run jobs much quicker than the older instances and that directly contributed to our success,” explained Mohan.

Mohan noted that the performance and reliability of the Intel processors in the various AWS instances have met and exceeded expectations since starting in the cloud: “With processors, execution speed is very important, which is why we chose the C5 and z1d instances, but reliability is also critical. Obviously, a glitch in the middle of a job that needs to run for three to four days is really problematic. In the year or so of our first design project, we experienced thousands of machines that were rock solid from an infrastructure standpoint with almost no crashes.”

Overall, the benefits of the cloud approach included faster turnaround times and faster iterations on designs and simulations. “We were able to create a higher quality chip much faster than would have been possible using our own resources. Doing what we did on premises would have been incredibly expensive and probably impossible.”



Jitendra Mohan, CEO of Astera Labs, explains the company uses the latest HPC instances on AWS to provide execution speed and reliability.

Solution Summary

By operating exclusively in the cloud, Astera Labs enjoys a high degree of operational flexibility. It not only can easily scale AWS compute resources up and down as needed to keep design processes speeding along and costs in check, it can also employ the resources that make the most sense for a given workload. AWS offers compute instances that include several generations of Intel processors and technology, including custom Intel Xeon Scalable processors and 2nd generation Intel Xeon Scalable processors. The ability to immediately take advantage of the latest Intel technology as new AWS compute instances become available will be an advantage for Astera Labs moving forward.

Where to Get More Information

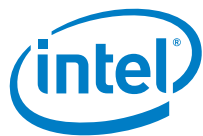
Learn more about [Intel high performance computing](#).

Learn more about [2nd generation Intel Xeon Scalable processors](#).

Learn more about [Astera Labs](#) products and solutions.

Learn more about [Six Nines IT](#) consulting offerings.

Find out about [HPC on AWS](#).



Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

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 www.intel.com/hpc.

Intel, the Intel logo and Xeon are trademarks of Intel Corporation.
Other names and brands may be claimed as the property of others.

© Intel Corporation