aws

carahsoft.

Astera Labs Partners with Exostellar to Realize a Scalable and Cost-Optimized AWS Infrastructure

Case Study

Thank you for downloading this AWS and Exostellar Case Study! Carahsoft is the distributor for AWS public sector solutions available via GSA, NASPO, The Quilt and other contract vehicles.

To learn how to take the next step toward acquiring AWS's solutions, please check out the following resources and information:

For additional resources:
carah.io/AWS-Resources

For upcoming events:
carah.io/AWS-Events

For additional AWS solutions:
carah.io/AWS-Solutions

For additional Marketplace solutions:
carah.io/AWS-Marketplace

To set up a meeting:
AWS@carahsoft.com
888-662-2724

To purchase, check out the contract vehicles available for procurement:
carah.io/AWS-Contracts

Astera Labs partners with Exostellar to realize a scalable and cost-optimized AWS infrastructure

EXECUTIVE SUMMARY

Astera Labs, a fabless semiconductor company, pioneered its Electronic Design Automation (EDA) workflow on Amazon Web Services (AWS) to help it scale compute requirements that can fluctuate by 10x during a project timeline. However, over the last several years, Astera Labs' AWS spend had increased significantly. To tackle the increased spending, Astera Labs partnered with Exostellar to optimize its design and verification workflows on AWS. During nine months of production deployment, Exostellar's X-Spot enabled Astera Labs to achieve significant cost reductions of up to 50% for on-demand compute resources while maintaining high performance and reliability.

CHALLENGE

During its chip design and verification process (Physical Design Workflow), Astera Labs relies heavily on EDA tools for the tape-out process. Tape-out is the final stage in chip design where a design is thoroughly verified and prepared for manufacturing. The tape-out stage is mission-critical and demands substantial compute resources and reliability. Any errors or oversights during this phase can lead to costly delays and manufacturing defects. To enhance performance and scalability, Astera Labs made the decision to migrate its workloads to the AWS cloud platform.

While running compute-intensive tape-outs with Synopsys and other tools using on-demand AWS instances, Astera Labs encountered the challenge of escalating compute costs. The unpredictable patterns of its workload made it difficult to optimize resource utilization effectively, resulting in inefficient spending and a significant increase in AWS expense.

RUNNING PHYSICAL DESIGN WORKFLOW USING X-SPOT ON AWS

To address significant spending, Astera Labs ran its HPC workloads in AWS using Exostellar's X-Spot, which is a state-of-the-art cloud optimization solution that allows users to seamlessly relocate stateful and compute-intensive EDA applications between spot instances and on-demand instances resulting in up to 5x increase in compute for the same dollar spent (80% cost reduction).

Astera Labs' Synopsys workloads were executed using X-Spot, which offers a reliable and consistent virtual environment for spot instances. These workloads were seamlessly scheduled within the existing Portable Batch System (PBS) job scheduler with minimal modifications to the job launch process.

> "Exostellar's X-Spot allowed us to leverage the cost-effectiveness of spot instances with the reliability as-if they were on-demand instances, allowing Astera to run more simulation and physical design workloads resulting in reduced development time and quicker time to market."

Jitendra Mohan
CEO, Astera Lab

AT A GLANCE

Challenges

- Significant increase in computing cost
- Inefficient spending
- Difficult to optimize resource utilization

Solution

- Execute electronic design automation workloads using X-Spot

Benefits

- 2x more compute for same cloud spend, i.e 50% cost reduction
- Eliminated the need for upfront capital investments and long-term commitments
- Scale easily based on fluctuating project demands

Astera Labs partners with Exostellar to realize a scalable and cost-optimized AWS infrastructure

exostellar

Additionally, X-Spot facilitated the automatic live migration of workers between on-demand and spot instances. Specifically, the on-demand instances utilized for the workloads were of type r6i, while the spot instances used were of r6/r5 families. This integration flow ensured optimal resource utilization and cost efficiency. Importantly, the live migration process between different instance types occurred smoothly, without changing network addresses or network connections allowing the job to continue running without any pausing or other interruptions.

Figure 1, Synopsys workload migration using X-Spot

PERFORMANCE AND RELIABILITY

Astera Labs has now increased compute capacity by up to 2x for the same budget to run thousands of chip design workloads by using X-Spot.

With X-Spot, Astera Labs was able to lower its escalating on-demand EC2 instances cost by 50% and achieve a scalable cost-optimized EDA infrastructure in full production. The significant cost savings were achieved by migrating mission-critical workloads 90,103 times between on- demand and spot instances reliably as shown in Figure 1 and Table 1.
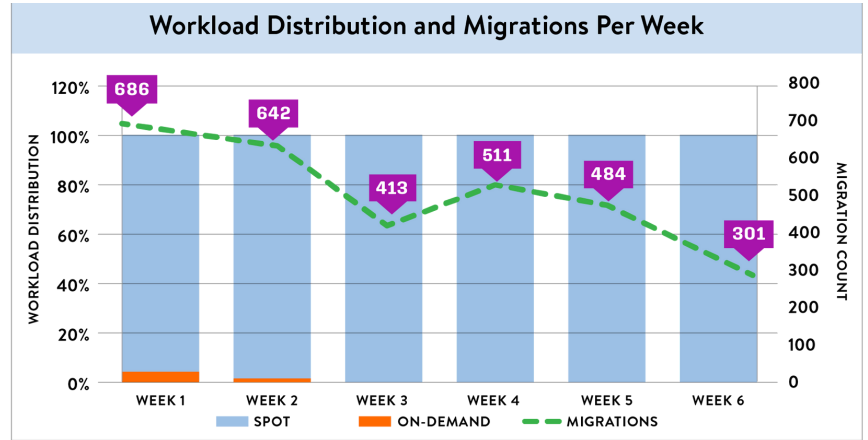
Workload Reliability Analysis Over 9+ Months of Deployment

Job Lifetime Analysis	TOTAL NUMBER OF JOBS	17,019
	AVERAGE JOB LIFETIME (DAYS)	0.7
	MAX JOB LIFETIME (DAYS)	30.0
	PERCENT OF TIME JOB IS ON SPOT INSTANCE	91%
	SUPRISE JOB TERMINATIONS	78
Spot Instance Lifetime Analysis	AVERAGE LIFETIME (HOURS)	3.4
	LONGEST LIFETIME (HOURS)	11.5
	NUMBER OF SPOT INSTANCES CREATED	140,490
Migration Analysis	% JOBS MIGRATED	55%
	% JOBS MIGRATED OVER 10 TIMES	16.9%
	MAX # OF MIGRATIONS FOR A SINGLE JOB	229

Table 1, X-Spot performance and reliability

SUMMARY

The adoption of Exostellar's X-Spot solution by Astera Labs has proven to be a transformative move with significant business implications, resulting in enhanced cash flow and decreased capital expenditure (CapEx) requirements.

Cost Reduction

Exostellar achieved cost reductions of up to 50% for on-demand compute resources which directly translates into lower expenses for Astera Labs, allowing the company to allocate its financial resources more efficiently. This also allowed Astera Labs to run more workloads in a shorter time period for the same cost as before, thus accelerating the chip development process and reducing time to market.

Operational Expenditure (OpEx) Model

Astera Labs pays for the cloud resources it uses on a consumption basis. This model eliminates the need for upfront capital investments and long-term commitments associated with buying compute savings plans from public cloud providers. By shifting to an OpEx model, Astera Labs can reduce its CapEx requirements and free up cash flow that would have otherwise been tied to infrastructure investments.

Scalability without Capital Investment

Astera Labs can scale its compute resources without the need for additional capital investments. X-Spot's ability to seamlessly migrate workloads between on-demand and spot instances enables Astera Labs to increase its compute capacity by up to 2x within its existing budget. This scalability allows Astera Labs to meet the fluctuating demands of its projects.