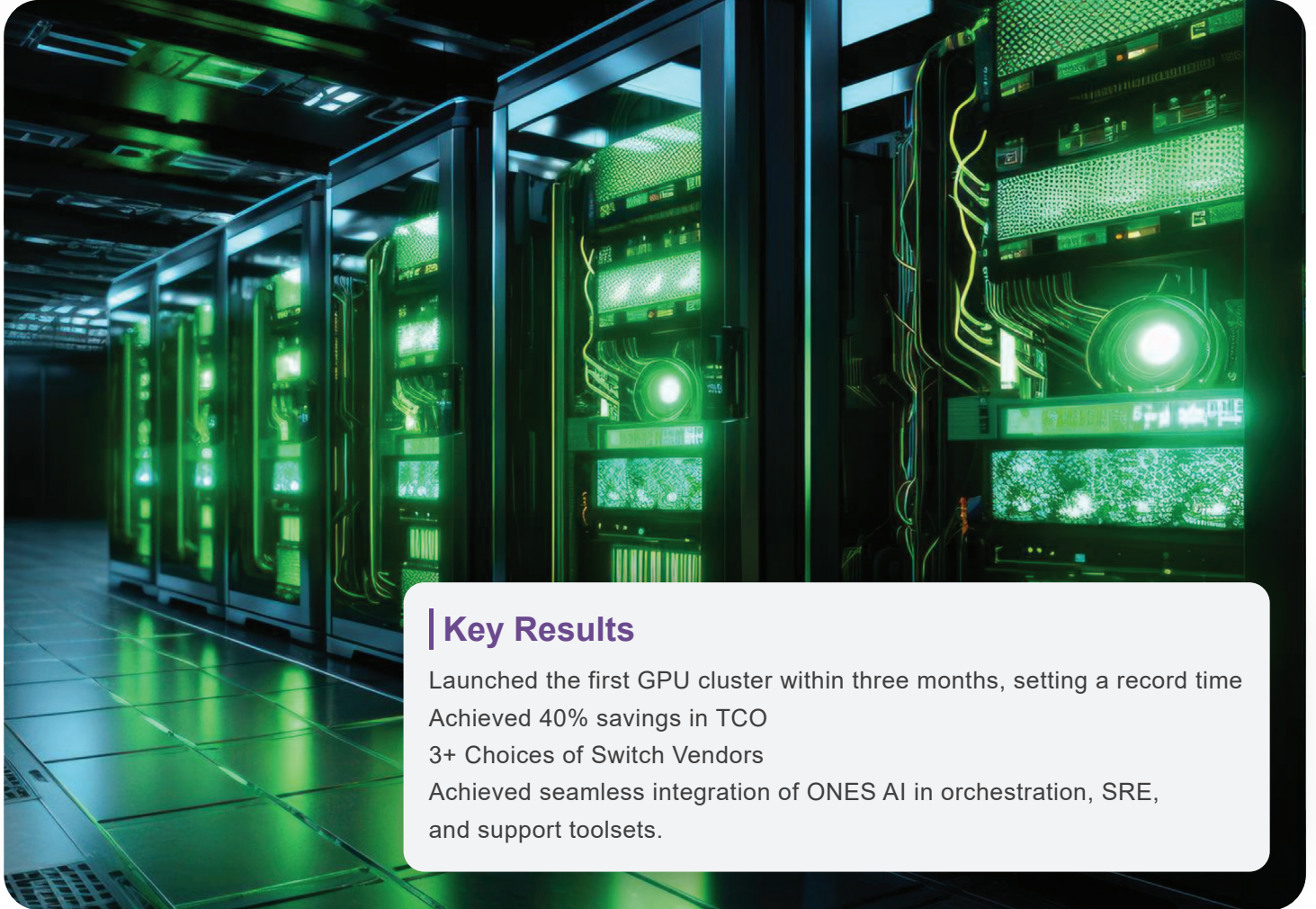




Case Study:

## Leveraging RoCE for Cost-Effective GPU Cloud Services



### | Key Results

- Launched the first GPU cluster within three months, setting a record time
- Achieved 40% savings in TCO
- 3+ Choices of Switch Vendors
- Achieved seamless integration of ONES AI in orchestration, SRE, and support toolsets.

## Overview

In the highly competitive arena of cloud-based GPU services, achieving top-tier performance without incurring substantial costs is a significant challenge. A leading cloud service provider was determined to enhance the performance of Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) for their GPU as a Service offerings. This objective was crucial for supporting high-performance computing (HPC) applications, including AI and machine learning workloads, in a cost-effective manner, without succumbing to the constraints and high costs associated with vendor lock-in, especially from proprietary solutions like NVIDIA's end-to-end offerings.

## Objective

The primary goal was to implement an RoCE solution that would leverage commoditized hardware components to deliver exceptional performance and flexibility. This needed to be achieved while avoiding the high costs and restrictive practices of established vendors, particularly NVIDIA, known for its pricey, proprietary solutions, and Arista, recognized for its premium networking hardware and software.

## Strategic Issue

The provider faced a dual challenge: competing with NVIDIA's high-end, integrated solutions and Arista's networking products, both of which come with significant cost implications and potential vendor lock-in. The market demanded a solution that offered both high performance and scalability at a lower cost, without being tied to the limitations of traditional, proprietary systems.

## Solution

After evaluating various alternatives, the company decided to implement the Aviz Open Network Enterprise Suite (ONES) for its GPU cloud as a service cluster.

Several compelling factors drove selection of Aviz as a vendor and ONES as the platform to manager and support their AI Fabric:

➤ **Vendor Agnostic:**

Aviz ONES stood out as a vendor-neutral solution, breaking the dependency on specific hardware manufacturers and enabling the company to leverage a broader range of equipment.

➤ **Expertise in SONiC:**

The suite's support for the SONiC platform, combined with the team's expertise, provided a robust foundation for RoCE, tailored specifically for AI-driven applications.

➤ **AI Fabric Design:**

The RoCE stack within Aviz ONES was explicitly designed for AI fabric, aligning perfectly with the company's needs and ensuring the highest TCO benefit.



➤ **Out of the box solution:**

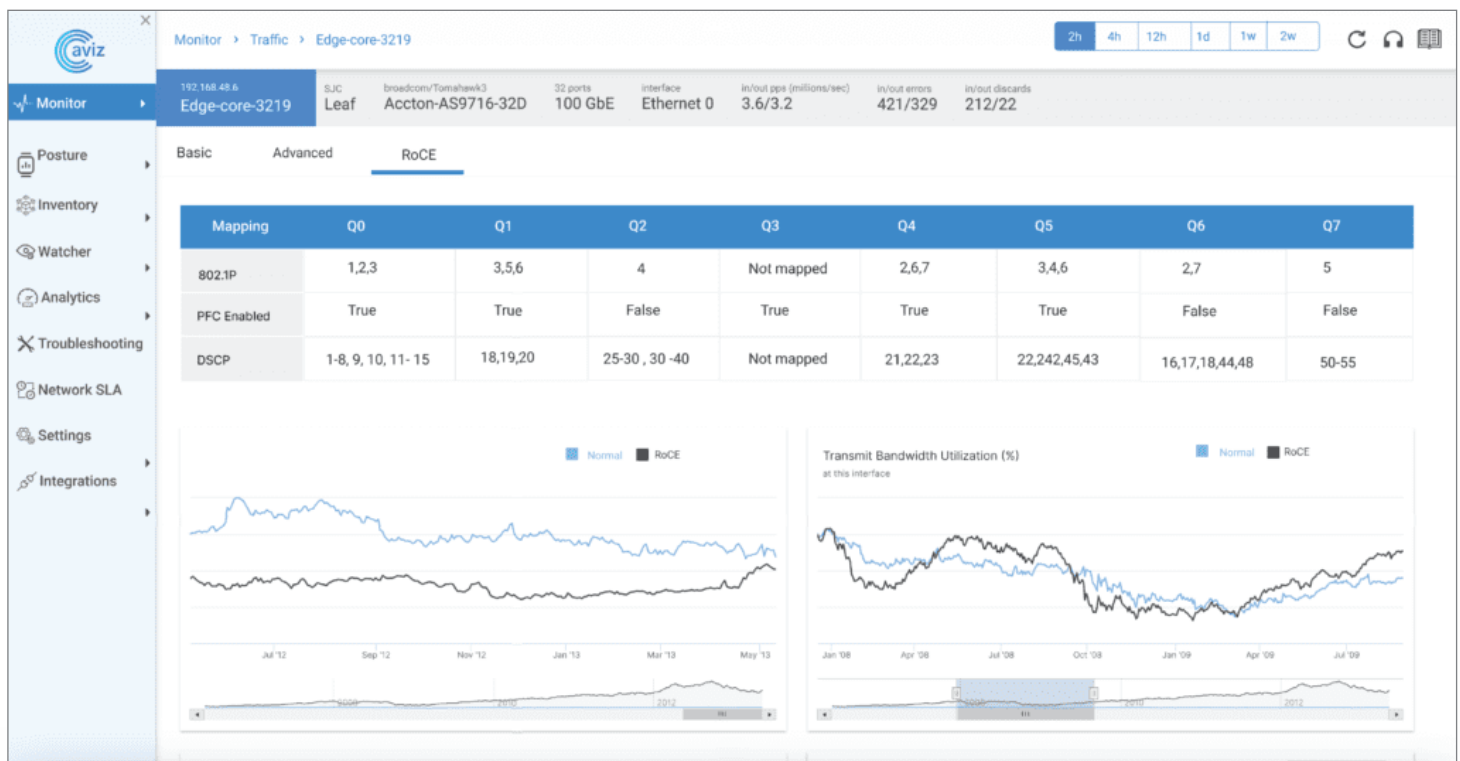
ONES SONiC recipes are quality tested and baked with various SONiC based solutions make it earlier for customers to deploy the these recipes right out of the box

➤ **24x7 Support for SONiC:**

ONES comes with SLA based 24x7 support which is first of its kind where a product not only provides the multi-vendor NetOps experience but also can help with the network issues with a 30 minute SLA.

➤ **Right price while enabling more choices:**

With ONES the customer got access to 35+ pre-tested SONiC solution which enabled customers to be future proof in their design, all while keeping the TCO for ONES without spiking the budget.



Implementation:

## Technical Specifications

**Speeds:** 100GbE and 400GbE

**Vendors:** Edgecore Switches and NVIDIA GPUs

**Features:** IP CLOS BGP, RoCE, ECMP and other Management features

## Competition

The main competitors in this space were identified as Arista and NVIDIA. Despite their strong offerings, these competitors were either too expensive or too restrictive in terms of vendor lock-in, making them less appealing for the company's specific requirements.

## Key Results

By adopting Aviz ONES, the cloud service provider significantly enhanced its GPU service offerings, achieving:

➤ **Open, Scalable Architecture:**

The solution avoided vendor lock-in, maintaining operational efficiency comparable to integrated solutions but with greater flexibility.

➤ **Competitive Edge:**

The provider managed to undercut the higher-cost solutions from NVIDIA and Arista, offering high-performance RoCE capabilities at a more attractive price point, thereby gaining a competitive advantage.

## Next Steps

Looking for a similar transformation? Contact Aviz Networks to discover how we can replicate this success for you. Visit us at [www.aviznetworks.com](http://www.aviznetworks.com).