



OVERVIEW

Six Nines designed and deployed a 'one-click' scripted solution for a leading virtual reality (VR) software and equipment provider in order to deploy an extremely complex, highly scalable High Performance Compute (HPC) environment on Amazon Web Services (AWS). The solution, which enables the company to perform advanced auditory testing in virtual reality simulations, utilizes Ansible and AWS CloudFormation to create a Slurm cluster on a group of 2 to 60 p2.16xlarge instances.

CHALLENGES

The VR company's team of research scientists is responsible for developing innovative sound propagation techniques for generating high-quality acoustics for VR and AR applications. While deeply technical in their areas of expertise, the customer's experience in building and scaling HPC clusters on AWS was limited. Additionally, the customer had previously run pre-existing clusters on Taito-CSC but found that they were resource-constrained. In order to run batch simulations at scale, the customer needed a team of AWS HPC experts to automate the rapid scale-up of compute resources.

SOLUTION

With almost ten years of experience designing and deploying HPC environments on AWS that scale beyond 100,000 cores, Six Nines' team of HPC experts were able to diagnose and resolve the situation accordingly. First, Six Nines consulted with the customer's research team to evaluate and identify the success factors. The customer needed to rapidly scale an ephemeral cluster with nearly 1,000 GPU resources, capable of utilizing up to 8TB of memory, to run tests against and store the results in a stable file system for later retrieval and review. Next, Six Nines designed and deployed the framework and scheduling mechanisms to validate the customer's code and dataflow. Six Nines used AWS CloudFormation Templates to deploy a secure Amazon Virtual Private Cloud (Amazon VPC) environment, a single Amazon Elastic Cloud Compute (Amazon EC2) node with Amazon Linux to run Slurm/OpenMPI as a scheduler and CUDA to utilize the GPU resources, and a bastion host for secure customer access. Following successful tests on a six-node P2 instance cluster with acceptable job outputs, Six Nines handed over the environment along with the Ansible and CloudFormation scripts to manage the MPI/CUDA configurations.

AWS SERVICES USED

- AWS CloudFormation
- Amazon EC2
- Amazon Simple Storage Service (Amazon S3)
- Amazon Linux
- Amazon VPC
- AWS Certified Advanced Networking
- Placement Groups
- Amazon EC2 Spot

SIX NINES EXPERTISE

- HPC
- AWS DevOps Competency
- AI/ML
- NVIDIA GPU

RESULTS

The customer was so pleased with the speed and improved capabilities of their cluster that they subsequently engaged Six Nines for additional larger projects in order to add functionality to the HPC solution and expand the solution globally. Leveraging AI/ML and new NVIDIA GPUs that AWS has made available, Six Nines is now helping the customer scale out so other disparate teams across multiple continents can use the resources in AWS to get to their research conclusions faster.

RESULTING CAPABILITIES

- Provide 1GB-10GB simulation instruction/input files distributed to multiple P2 nodes
- Provide 1TB-8TB memory per simulation
- Repeatedly execute 10,000 to 100,000 time steps per node, coordinate results across the nodes via MPI, and then iterate over this process before generating a results file
- Launch clusters up to 60 P2 Nodes and nearly 1,000 GPU
- Rapidly deploy, scale, and terminate multi-node HPC compute processing in short intervals (~30 minutes)



Premier
Consulting
Partner

- DevOps Competency
- Microsoft Workloads Competency
- Solution Provider

WHY SIX NINES?

Six Nines IT is an AWS Partner Network (APN) Premier Consulting Partner and AWS Solution Provider specializing in helping businesses migrate to the cloud responsibly. A member of the APN since its inception, Six Nines has successfully migrated hundreds of customers across all industries to the cloud and offers an unparalleled combination of speed, agility, experience, and proprietary solutions to deliver accelerated solutions and a rapid time-to-value. The Oakland-based company combines old-school, on-premises IT roots together with deep expertise and a laser focus on all things AWS – including a core concentration on High Performance Computing and Microsoft Workloads (AWS Microsoft Workload Competency and AWS DevOps Competency) – to deliver bespoke solutions that are individually tailored to meet customers’ unique needs throughout the cloud lifecycle.

