

SUCCESS STORY | ZENUITY

SCALABLE AI INFRASTRUCTURE ACCELERATES AUTONOMOUS VEHICLE DEVELOPMENT



Image courtesy of Zenuity



Zenuity deploys scalable AI platform from NVIDIA® with Pure Storage® to accelerate development of smarter, safer autonomous vehicles

DEVELOPING THE VEHICLES OF THE FUTURE



Image courtesy of Zenuity

Autonomous vehicles will forever change the \$10 trillion transportation industry, making roads safer, cities more efficient, and transforming how we travel and transport goods. One day, every car will be a self-driving car. As companies race to develop autonomous vehicles (AV) they are facing huge challenges.

At the core of successfully developing autonomous vehicles (AV) is the software running on powerful AI car computers that will be entrusted with making crucial, real-time decisions. Deep learning is enabling these AV systems to result in dramatically safer roadways.

Zenuity, a Sweden-based company with offices also in Germany and the United States, is tackling the greatest challenges facing the automotive industry by developing a software strategy that places the company's expertise in artificial intelligence (AI) and deep learning at the center. A joint venture of Volvo Cars and Veoneer (formerly Autoliv), Zenuity will develop the software 'brain' of the car built on the while Veoneer will bring the technology to the global market.

CUSTOMER PROFILE



Organization
Zenuity

Industry
Automotive

Location
Sweden, Germany, USA

Website
www.zenuity.com



Image courtesy of Zenuity

PRODUCTS

NVIDIA® DGX-1™

Pure Storage® FlashBlade™

Tensorflow

PyTorch

ZENUITY USES DGX-1 AND FLASHBLADE TO ACCELERATE AUTONOMOUS DRIVING

The deep learning work at Zenuity has its roots in a project that began at Autoliv in 2015. Benny Nilsson, who is now Manager of Deep Learning at Zenuity, was part of that original project. “We spent a lot of time at Autoliv debating and testing for the best architecture for deep learning,” Nilsson recalled. “Our first system was an off-the-shelf server with GPUs (graphics processing units) and a typical enterprise storage system. We learned a lot from that system, both in terms of what works and what doesn’t.”

This prior experience with GPU technology proved invaluable when thinking about what an optimal deep learning training platform needed to deliver at Zenuity. They knew they needed ground-breaking performance to handle the exponentially increasing volume of sensory data to train their models. This would permit them to iterate on models developed rapidly without losing valuable development cycles. Additionally, their researchers couldn’t afford to waste time playing the role of IT system integrators; designing, building and troubleshooting hardware and software or scouring open source forums in search of answers.

One compute platform emerged as the best fit for their needs — the DGX-1 deep learning platform from NVIDIA. “DGX-1 is clearly the gold standard when it comes to AI and deep learning,” Nilsson noted. DGX-1 is engineered to deliver the highest deep learning training performance while offering the fastest deployment model with a plug-and-play deployment experience, and simplicity of management and day-to-day operations for data scientists.

“We are in a global race to see who can deliver the safest solution for autonomous vehicles first. The combination of NVIDIA DGX-1 and Pure Storage FlashBlade provides a breakthrough AI infrastructure that enables us to take a leadership position.”

Benny Nilsson,
Manager of
Deep Learning, Zenuity

Having decided upon a compute platform, the team turned to storage. “In our first system, we didn’t fully understand the importance of storage,” Nilsson said. “When we started testing the DGX-1, the storage system we had could not feed the GPUs with enough data to keep them busy. Data was the bottleneck and that’s when we realized we needed a modern storage system.”

Ola Eriksson, infrastructure expert at Zenuity, oversaw the selection process for storage. He identified ten storage vendors, and then developed a rigorous test regimen representing real-world deep learning workloads to compare their performance. “Our top priority was to ensure that all the DGX-1s were kept busy,” Eriksson said.

Added Samuel Scheidegger, Research Engineer on the deep learning team, “DGX-1 utilization directly translates to data scientist productivity. So before we purchased any storage system, we had to be able to test it in a hands-on environment.”

Eriksson said the key criterion was the storage cost per DGX-1 that Zenuity would rely on to provide sufficient bandwidth for deep learning workloads. “We evaluated how many DGX-1s each storage solution could deliver data to without throttling as the basis for calculating storage costs. On this measure, FlashBlade from Pure Storage was the clear winner.”

FlashBlade is architected with a modern, massively parallel architecture, eliminating serial bottlenecks that hold legacy storage systems back. It is engineered to deliver the performance essential for AI and deep learning while offering simplicity, so data scientists can focus on their data, not the infrastructure.

“With its compact form factor and simplicity, FlashBlade delivers a petabyte in just 8U,” Scheidegger observed.

DELIVERING THE POWER TO MAKE CARS SMARTER AND SAFER

Training as well as testing deep neural networks require a huge amount of computational power and tremendous amounts of data. In the development of safety critical systems used in self-driving cars, detection accuracy requirements are at the highest level. As such, the computational demands placed on Zenuity’s IT infrastructure are massive.

“By combining NVIDIA DGX-1 with Pure Storage FlashBlade our data scientists can enjoy the best of both worlds — processing the amount of data needed to ensure safety without compromising product timelines.”

Benny Nilsson,
Manager of
Deep Learning, Zenuity

The primary workloads are neural network training jobs where GPUs are processing huge amounts of data gathered from test cars. The data consists of images from surround cameras, as well as from radar and LIDAR(a laser scanner). Data scientists at Zenuity utilize the most popular frameworks being TensorFlow and PyTorch.

By processing increasing volumes of data with greater frequency, the autonomous driving software continues to learn and become more accurate over time. But these continuous iterations of network training are extremely time-consuming. Datasets for each training run can reach into the terabytes, and a single network training run can take weeks if the environment is not properly optimized. Zenuity is tackling this with the combined performance of DGX-1 and FlashBlade, which is enabling them to make ground-breaking progress by reducing training run intervals as they iterate on their models.

“The industry constantly faces a trade-off,” Nilsson said, “between the need for processing more data — which is essential to deliver ever-greater improvements in safety — and the desire to accelerate time-to-market. By combining NVIDIA DGX-1 with Pure Storage FlashBlade our data scientists can enjoy the best of both worlds — processing the amount of data needed to ensure safety without compromising product timelines.”

Zenuity currently has two FlashBlade systems —each equipped with fifteen 52TB blades— connected to multiple DGX-1s. Each DGX-1 includes eight Tesla V100 GPUs —the world’s most advanced data center GPU— which are configured in an NVLink based hybrid cube-mesh topology for incredibly fast inter-GPU communications. FlashBlade keeps all DGX-1s busy with data, with significant performance headroom to give the Zenuity team the assurance that the combination of NVIDIA and Pure technology will meet their needs well into the future.

“Our training data sets are already in the hundreds of terabytes, and that number will increase exponentially as we collect more data and put more cars on the road,” Nilsson said. “With the limitless scalability of the DGX-1 and the FlashBlade system, our deep learning infrastructure is future-proof, able to scale in capacity by simply adding more storage blades to ensure we’re feeding enough data to keep the DGX-1s busy.”

ACCELERATING INTO THE FUTURE

The AI infrastructure of DGX-1 and FlashBlade has met Zenuity's number-one priority — keeping its IT investment fully utilized. “If you're not able to feed the GPUs with the data they need, then you are underutilizing your most valuable resources — data scientists,” Nilsson said. “The combination of NVIDIA DGX-1 and Pure Storage FlashBlade is helping us to reduce training times, which means IT investment is better utilized and data scientists are more productive and, ultimately, happier.”

For Zenuity, having systems which are simple to set up and easy to deploy has also been a crucial factor. “The DGX-1 and FlashBlade, are simple to set up and require very little attention from day to day,” Eriksson said. “They both deploy quickly and you need very little support thereafter.”

Over time, Zenuity expects to grow its DGX-1 environment. This will be in direct response to the increasing size of their datasets; the need to tap into the power of multi-GPU training to process that data; and mounting business pressure to iterate faster and develop new, smarter autonomous driving neural network models. As Zenuity grows its DGX-1 based AI platform, the company's broader infrastructure will need to keep pace, especially when it comes to storage.

“We are in a global race to see who can deliver the safest solution for autonomous vehicles first,” Nilsson noted. “The combination of NVIDIA DGX-1 and Pure Storage FlashBlade provides a breakthrough AI infrastructure that enables us to take a leadership position.”

www.nvidia.com



© 2018 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, NVIDIA DGX-1, and NVIDIA Tesla are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. JUN18

