



INSIGHT AND ACTION

Future of Sustainable Data Centers

Penguin Solutions, NVIDIA, and Shell are
Reimagining the Future of Sustainable Data Centers



Introduction

In recent years, the computing landscape has undergone many changes.

The world is increasingly reliant on high-performance computing (HPC), artificial intelligence (AI), and machine learning (ML) to power everything from weather forecasting to self-driving cars. These technologies have enabled us to solve previously unsolvable problems, make discoveries, and create experiences that were once impossible.

Despite the advantages of these technologies, the amount of energy they consume presents a challenge. HPC, AI, and ML place an enormous burden on traditional data centers. According to the International Energy Association, in 2021, data centers accounted for 0.9-1.3% of global electricity consumption. HPC and AI applications are particularly demanding due to their compute-intensive workloads. The stage is set for a new computing era defined by faster speeds, more efficient data centers, and a greater focus on sustainability.

“The Penguin Solutions team is pleased to be collaborating with Shell and NVIDIA to deliver the next generation of bespoke

data center cooling solutions that are placing sustainability at the forefront. We’re proud to be supporting Shell’s transition of their business to net-zero emissions and other customer-centric sustainability efforts while providing the advanced computing infrastructure necessary to handle such intensive HPC, AI, and ML workloads. We look forward to pushing the innovative boundaries with Shell and NVIDIA as together we lead the way for a more sustainable future.”

Kevin Tubbs, Vice President of Strategic Solutions Group at Penguin Computing.

At Penguin Solutions, we empower the world’s most innovative organizations to push the boundaries of possibility with advanced computing infrastructure that supports AI and ML applications. As a leading provider of high-performance computing solutions, we collaborate with some of the world’s largest organizations. That’s why we are working with Shell and NVIDIA to harness the power of technology to deliver game-changing sustainability benefits with bespoke data center cooling solutions.

- ▶ Shell offers a range of renewable power and energy efficiency solutions to help make data centers more sustainable, including liquid immersion cooling technology.

- ▶ NVIDIA HGX A100 is a scalable, high-performance server platform that supports AI and HPC applications and is ideal for use in data centers.
- ▶ Penguin Solutions brings together the combined power of innovative technology to create custom solutions that meet our customers’ business needs and goals.

United in a shared belief that sustainable data centers are the future, Penguin Solutions, NVIDIA, and Shell are unlocking new possibilities in modern data centers’ performance, efficiency, and environmental footprint.



“As digitalization across virtually all parts of society drives rapid growth in data centers, the need to manage the associated environmental impacts grows with it. We believe immersion cooling technology is an essential piece of the puzzle in tackling data center energy use. Shell Immersion Cooling Fluids can help data center operators cut costs, boost performance and achieve their sustainability goals, and deliver the greatest benefits when integrated with renewable power and other smart energy management solutions.”

- **Sjors van de Rijt**, General Manager Tech Sector, Renewables and Energy Solutions, Shell

The Challenge

Improve data center sustainability without compromising performance.

High-performance data centers demand greater power density, generate more heat, and require more cooling than traditional, air-cooled centers. As demand for HPC, AI, and ML increases, so does the need for more servers which means more space and power consumption.

In 2021 alone, global internet traffic grew by 39%. This growth translates into rising costs for data center operators and increasing CO2 emissions. Therefore, data center operators and energy suppliers must continue to find new ways to operate sustainably and mitigate environmental impact.

The Opportunity

Develop solutions that are good for both business and the environment.

Sustainability is high on the Shell agenda, with a range of near-term targets to support its goal to become a net-zero emissions energy business by 2050. The company is taking action today to reduce the environmental footprint of its operations and the energy products it sells.

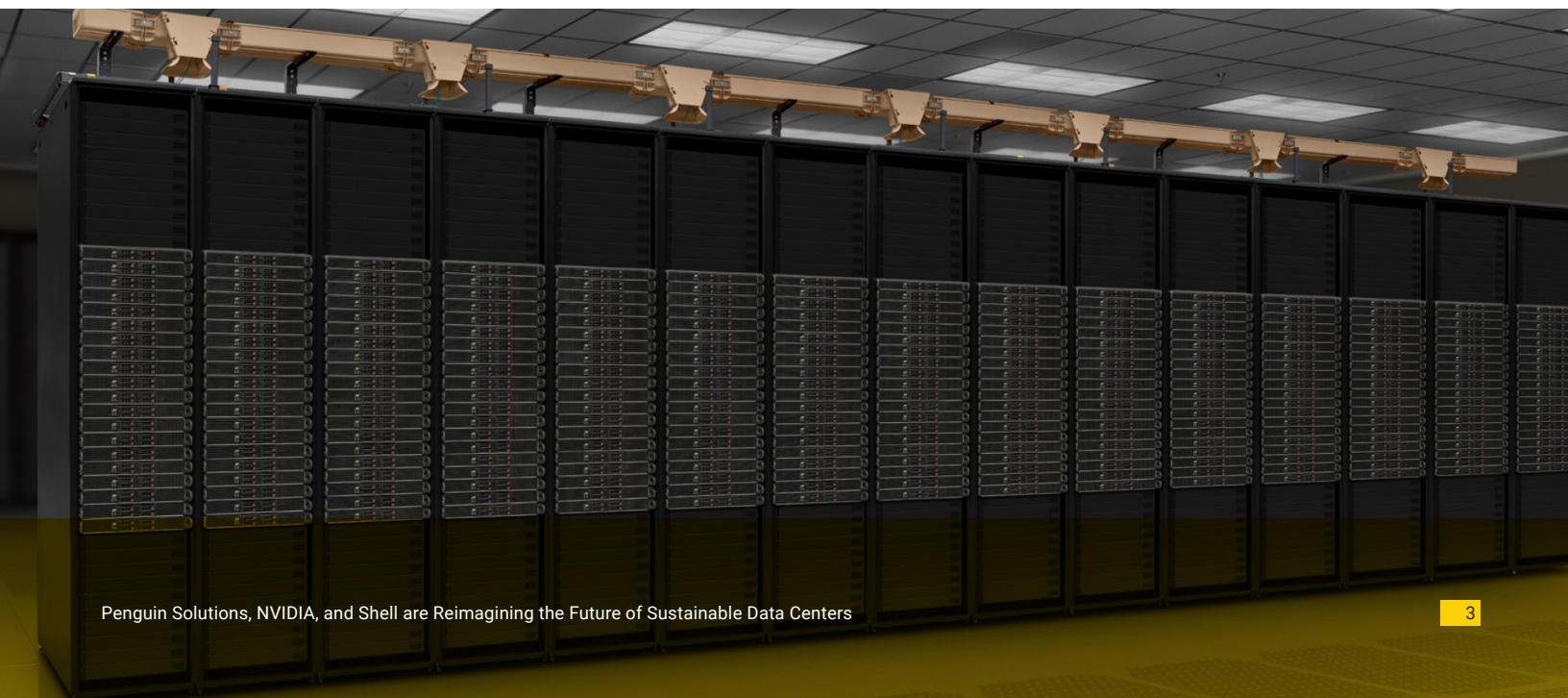
Data center operators are under increasing pressure to reduce the environmental footprint of their operations while remaining cost and performance-competitive. Shell views liquid immersion cooling technology as a critical component of an integrated solution to help make data centers more sustainable, both in its own HPC clusters, and in support of its customers' climate action ambitions.

The Technology

Engineering for exceptional performance with clear sustainability benefits.

Liquid immersion cooling involves submerging server components in a thermally conductive, electrically non-conductive (dielectric) coolant, which transfers heat away from the components more effectively than air. With immersion cooling, data center operators can significantly reduce the need to install server cooling hardware.

Immersion cooling also allows for much higher server densities than air cooling. As a result, more servers can be packed into a smaller space, leading to lower energy consumption, a smaller carbon footprint, and reduced operating costs.



Shell Immersion Cooling Fluids

HPC and artificial intelligence (AI) are two of the most challenging applications for data centers. They require immense amounts of processing power and generate a lot of heat. That's where Shell's immersion cooling fluids come in.

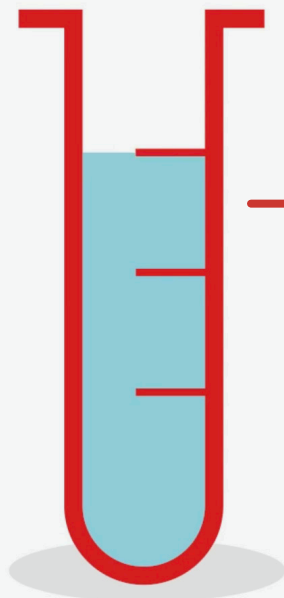
Shell Immersion Cooling Fluids are specifically developed for use in single-phase immersion cooling systems. Single-phase immersion cooling is the most cost-effective and sustainable method for cooling server banks, especially high-density racks of 50 kW or more.

Leveraging nearly half a century of expertise in fluid formulation and based on Shell's proprietary gas-to-liquids (GTL) technology, Shell Immersion Cooling Fluids maximize the energy efficiency of data servers and IT components across the most demanding of applications.

Shell Immersion Cooling Fluids are inherently stable and provide superior performance and material compatibility with server components. The synthetic fluids are clear, odorless, safe, and easy to handle.

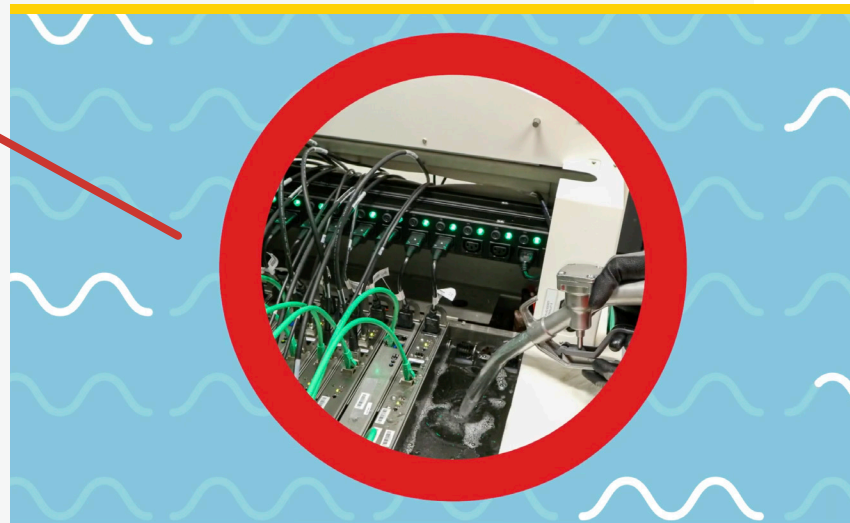
The high cooling efficiency, flow behavior, and excellent thermodynamic properties of Shell Immersion Cooling Fluids mean considerably less energy is required to run a network than traditional air cooling. This can translate to lower energy costs and emissions while improving server performance and location flexibility.

This fluid technology is backed by extensive component-level material compatibility testing and thermal simulation at Shell's Technology Centers, a robust quality assurance process, and one of the world's largest global supply chains.



Shell Immersion Cooling Fluid

A synthetic, dielectric, single-phase immersion cooling liquid made from natural gas, using Shell's Gas-To-Liquid technology



NVIDIA HGX A100

Massive datasets, exploding model sizes, and complex simulations require multiple GPUs with fast interconnections and a fully accelerated software stack. The NVIDIA HGX™ AI supercomputing platform combines the full power of NVIDIA GPUs, NVIDIA® NVLink®, NVIDIA InfiniBand networking, and a fully optimized NVIDIA AI and HPC software stack from the NVIDIA NGC™ catalog to provide the highest application performance.

NVIDIA is best known for its powerful graphics processing units (GPUs), used in gaming computers and high-end laptops. However, GPUs are extremely popular for ML and AI applications since they can provide massive computational power while remaining energy-efficient.

The HGX A100 combines NVIDIA Ampere architecture-based A100 Tensor Core GPUs with NVIDIA Quantum InfiniBand networking to provide up to 2 petaflops of computing power in a 4U chassis. With its ability to scale up to 64 GPUs, the HGX A100 is ideal for large-scale HPC applications such as deep learning, medical imaging, and financial modeling.

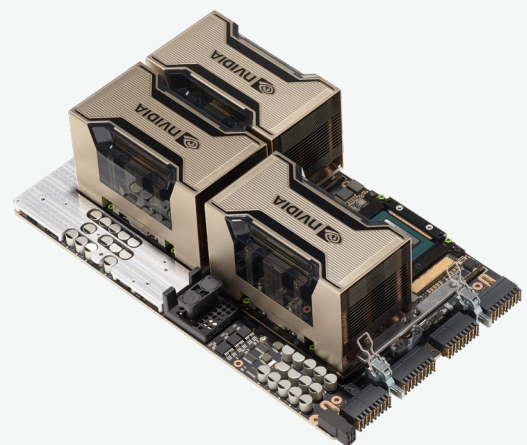
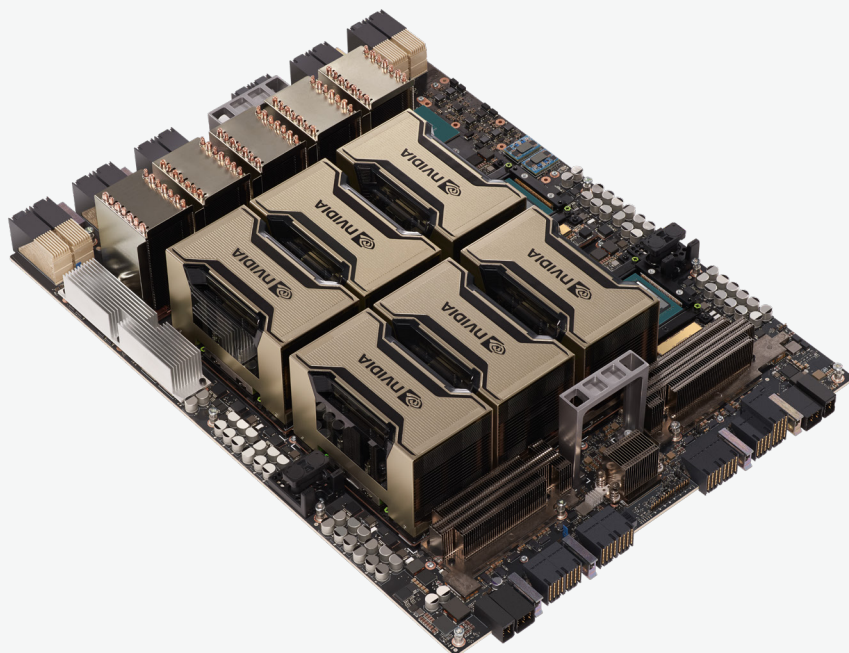


“As opposed to submitting a job to a CPU that may have 60 cores, you can send the same job to 10,000 cores. It’s a fantastic solution for applications that require immense processing power like high-performance computing and AI workflows.”

– **Marc Spieler**, Global Business Development - Energy Industry, NVIDIA

In addition to its raw computational power, the HGX A100 also features NVIDIA NVLink technology, which gives each GPU direct access to the others’ memory, eliminating the need for data movement between the GPUs. This reduces latency and improves performance.

The HGX A100 also includes NVIDIA Tensor Cores, which are purpose-built for deep learning workloads. Combined with its high-bandwidth networking and scalability, the HGX A100 is ideal for data centers looking to deploy HPC applications.



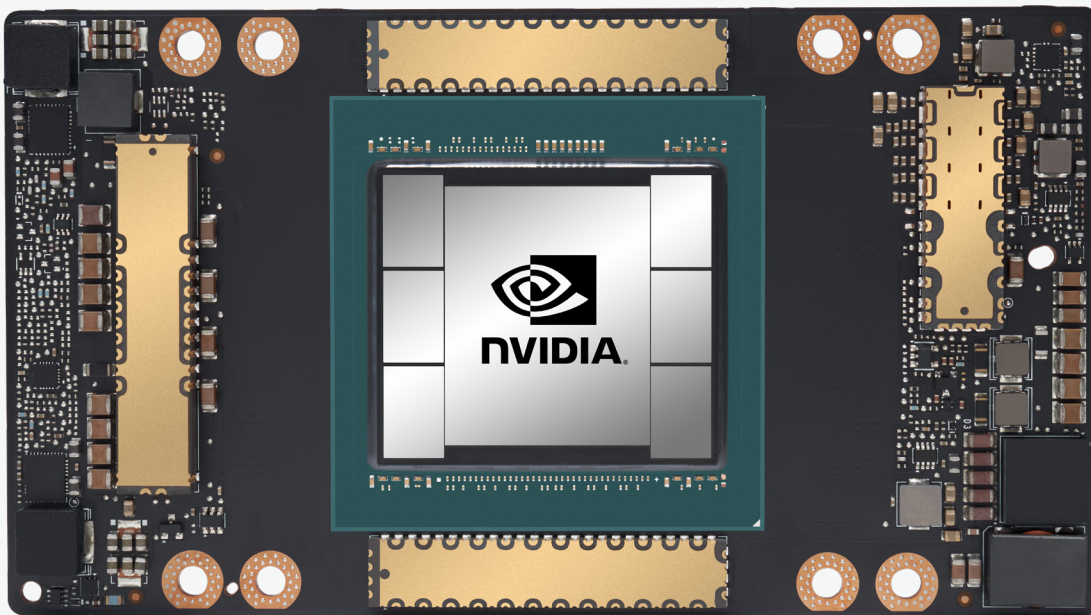
Penguin OCP-inspired Tundra Systems

Penguin Solutions has been a leading HPC solution provider for over 20 years. We offer some of the most energy-efficient supercomputing infrastructures in the world. We were one of the first companies to adopt liquid cooling for our systems, which can result in up to 50% energy savings compared to traditional air cooling methods

Penguin Solutions's Open Compute Platform (OCP) inspired systems can be used for single-phase immersion cooling and benefit from NVIDIA GPUs and Shell's Immersion Cooling Fluid.

Penguin's OCP-inspired servers are compatible with various industry-standard immersion cooling fluids, including Shell's Immersion Cooling Fluid.

In addition to OCP-adapted servers, Penguin also offers NVIDIA GPU-powered servers that are perfect for use in immersion-cooled data centers. NVIDIA GPUs provide unparalleled performance for demanding applications such as deep learning and video streaming. Combined with the high thermal conductivity of Shell's Immersion Cooling Fluid, Penguin's NVIDIA GPU-powered servers can offer exceptional performance while operating at lower temperatures than traditional air-cooled systems.



The Journey

Pushing the boundaries of possibility together.

In 2017, Penguin Solutions, NVIDIA, and Shell first partnered to upgrade Shell's HPC capacity in Amsterdam augmenting traditional CPU-based architecture with new GPU-powered capabilities from NVIDIA. This resulted in improved energy

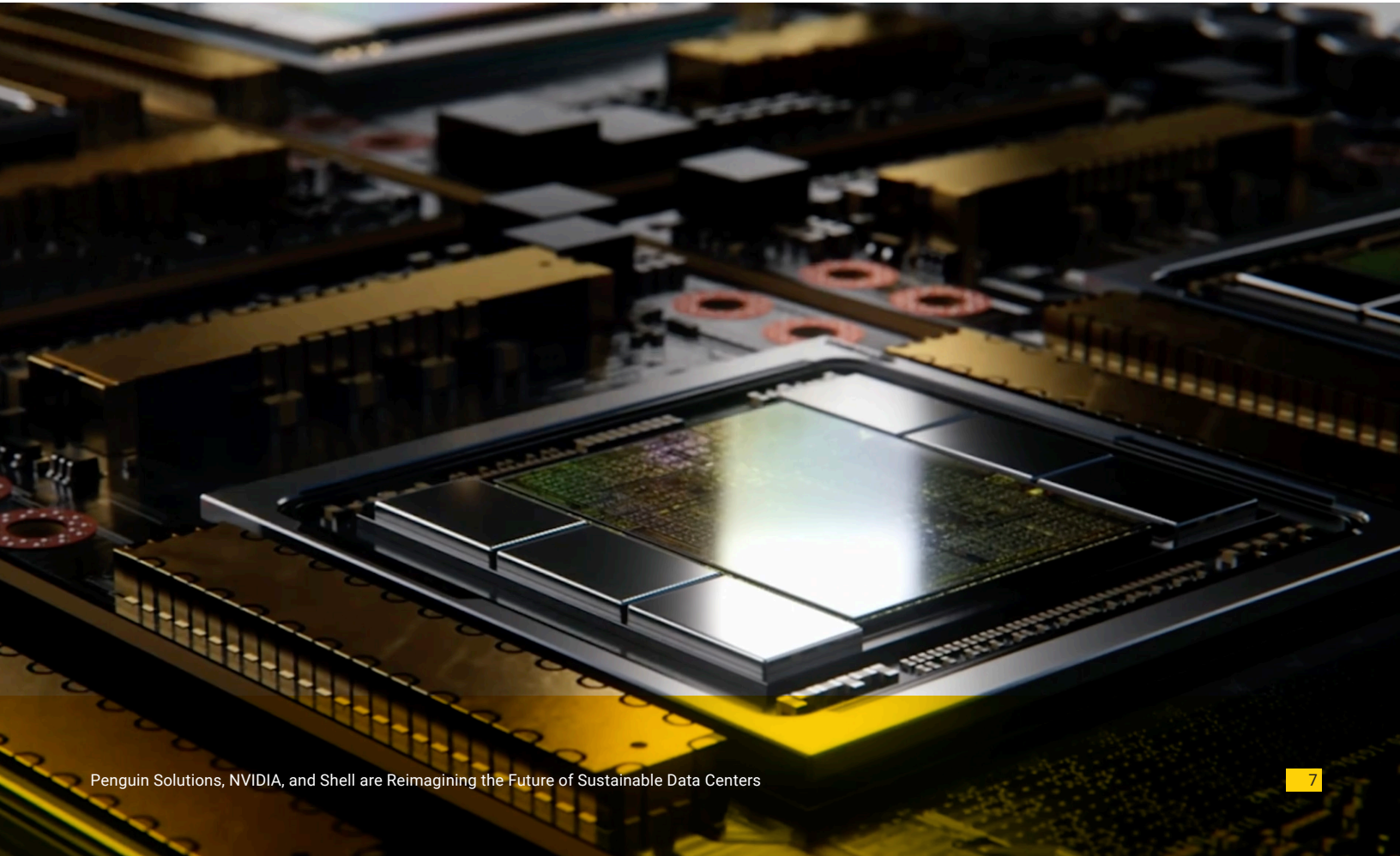
efficiency and increased computing capacity at the facility.

Penguin next implemented direct-to-chip cooled computing with Penguin OCP systems and NVIDIA HGX™ GPUs in Shell's data center in Houston, TX. The Houston facility integrates 100% renewable power and is designed to allow for future liquid cooling of the servers.

Back in the Netherlands, the collaboration has only grown stronger. The trio's

combined expertise most recently helped transform Shell's HPC cluster within a T-Systems-managed Amsterdam data center to a single-phase immersion cooling system.

Penguin Solutions adapted its OCP systems to fit Shell's needs. These systems were paired with NVIDIA GPUs to create a powerful, energy-efficient solution. The entire system was cooled using Shell Immersion Cooling Fluid.



Results

Reimagining the future of sustainable data centers

By teaming up with Penguin Solutions and NVIDIA, Shell saw increased energy efficiency and greater capacity to handle compute-intensive workloads in its Houston data center.

The integrated cooling solution implemented in Shell's HPC cluster in Amsterdam is helping Shell deliver high-end GPU processing power while reducing its energy consumption within the data center.



"The Amsterdam cluster upgrade, which is reducing the data center carbon footprint and operational costs, demonstrates a fully integrated, optimized, and scalable solution suitable for enterprise-level HPC. Continued adoption of immersion cooling is key to our partners' and customers' sustainability goals. Through our work with Penguin Solutions and NVIDIA, we are pleased to be able to build the positive case for this critical technology."

— **David Baldwin**, HPC Program Manager, Shell

The success in Amsterdam has spurred the trio to accelerate the transition to a fully immersed system in Houston

at Shell's Houston data center in the coming year.

The companies' collective aim is to help data centers operate more sustainably by promoting renewable energy sources paired with technologies that optimize energy use.

With the limitations of traditional cooling methods, Shell believes it's only a matter of time before immersion cooling becomes the standard in future data centers.

Penguin Solutions, Shell, and NVIDIA are excited to be at the forefront of adopting this technology. The trio believes this integrated solution has the potential to enable game-changing results in improving data center sustainability.

Learn More

To learn more about Penguin Computing, visit: penguinsolutions.com/computing/

To learn more about Shell's Immersion Cooling Fluid, visit: www.shell.com/immersion

To learn more about NVIDIA solutions for energy-efficient data centers, visit: NVIDIA.com/en-us/data-center/

