**BUILDING ENERGY EFFICIENT AI COMPUTING**

AS THE ENERGY DEMANDS OF AI GPUS CONTINUE TO RISE, WHAT STRATEGIES CAN IMPROVE THE ENERGY EFFICIENCY OF AI WORKLOADS IN DATA CENTERS?

**JAAP ZUIDERVELD**
**NVIDIA**

# Nvidia's Perspective on Energy-Efficient AI Workloads for Data Centers

December 1, 2024

Concerns regarding the power consumption of AI are widespread. Therefore, it was intriguing to gain insights from Nvidia, the leading chip manufacturer. Their Blackwell GPU, which is 25 times faster than its predecessor, will require a 150-kilowatt rack, whereas a standard rack typically consumes between 10 to 50 kilowatts.

However, executing a workload on 1,000 CPU servers can be accomplished with just 2 GPU servers, yielding energy savings of 80 to 90%.



NVIDIA'S ACCELERATED COMPUTING SOLUTION TARGETS AN ENERGY REDUCTION OF 80-90% BY OPTIMIZING THE ENTIRE DATA CENTER ENVIRONMENT.
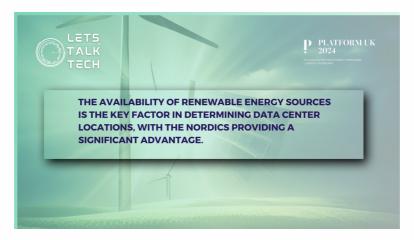
Accelerated computing energy savings

Nvidia successfully implemented Elon Musk's $3 billion AI factory in just

eight weeks, deploying a team of 300 engineers instead of the usual 1.5 to 2 years. This factory represents the largest compute environment to date, featuring intricate cabling that connects 100,000 GPUs to operate as a cohesive unit.



The $1 trillion sustainable compute opportunity

The transformation of data centers with AI is so significant that a CEO of a European data center mentioned that transitioning a legacy data center to AI is impractical; it's best to start anew.



The data center location trend of powered land with renewable energy

We are spotlighting thought leaders part of our annual "2024 Insights 2025 Trends" series while simultaneously soft launching our updated storytelling format designed to improve the clarity of integrated cross industry knowledge. Your thoughts, comments, requests and queries are always welcome as we continue to delve into the rapid evolution of the AI and energy infrastructure globally.

You and 10 others ·     3 reposts