



Power the next generation of intelligent, resilient energy systems on AWS and NVIDIA

Learn how power and utilities companies can boost application performance and transform energy operations with high performance computing (HPC) and AI across generation, transmission, and distribution.



Table of contents

Introduction	3
Key issues impacting power and utilities companies	4
Power and utilities use cases	5
Solutions	6
Customer stories	7
The AWS and NVIDIA advantage	8
How to get started	9

Innovating for energy efficiency

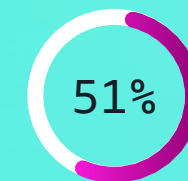
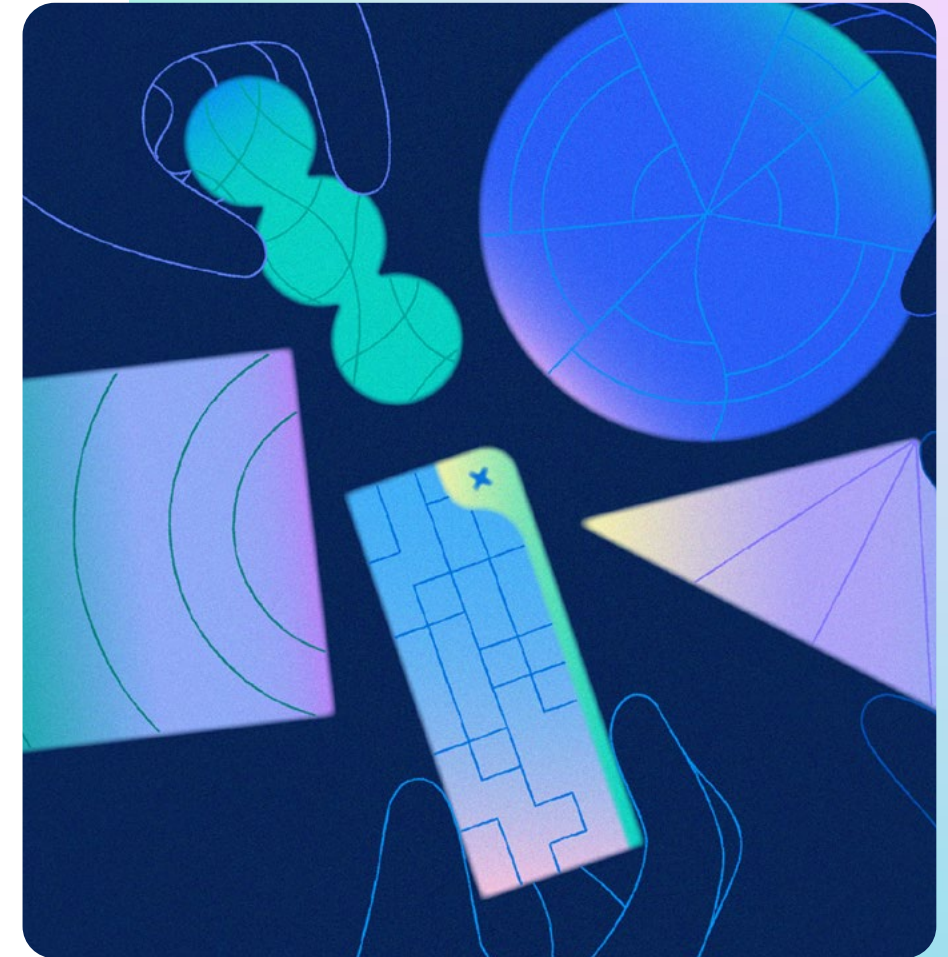
Power and utilities companies are accelerating digital transformation to modernize aging infrastructure, enhance grid reliability, and meet growing energy demands.

As they transition to more data-driven operations, HPC and AI are becoming essential for optimizing generation, transmission, and distribution. Traditional on-premises infrastructure has its limitations particularly as utilities require greater compute capacity to process massive datasets and model complex grid simulations.

As the demand for energy grows, the need to adapt to emerging technologies increases. Power and utilities companies are under pressure to modernize aging infrastructure, transition to low-carbon or renewable energy systems, and navigate climate and policy trends. At the same time, they must cut billions in maintenance costs from unplanned downtime, explore ways to reduce customer energy bills, ensure operational efficiency, and maintain control over sensitive energy data while meeting strict regulatory compliance.

AI, including gen AI, has the potential to boost electricity production, improve operational efficiency, reduce energy costs, and increase reliability. According to Deloitte's AI for Energy Systems report, AI integration could deliver more than \$200 billion of annual cost savings by 2030, and almost \$500 billion by 2050.¹ Furthermore, 51 percent of energy leaders surveyed in the Utility Innovation Survey 2025 said the technology will have the greatest impact on innovation in the next two years.²

To meet future energy demands for homes and businesses, power and utilities companies need secure, accelerated computing solutions that deliver scalability, domain-specific workload acceleration, greater software agility, and optimized cost efficiency all while having complete control over their data.



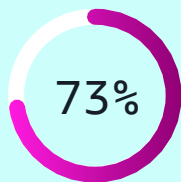
of energy leaders say AI will have the greatest impact on innovation in the next two years

Key issues impacting power and utilities companies

Power and utilities face growing pressure to monitor grid health in real time to ensure reliability, cut unplanned maintenance costs, and explore ways to lower energy bills amid rising rates. Like the oil and gas sector, they deal with aging grid infrastructure and labor-intensive upkeep, especially in transmission and distribution.

Traditional grids weren't designed for bi-directional energy flow from renewable sources. In addition, climate change and extreme weather adds further risk and demands for better prediction of outages and grid impact.

Power and utilities providers are modernizing rapidly to integrate renewables and strengthen grid reliability. According to Ernst & Young, 73 percent of power and utility companies are already using AI—including gen AI and ML—and are in the process of developing improvements or new uses.³ Yet delivering affordable, reliable, and sustainable energy remains among the industry's major challenges.



of power and utility companies use AI, including gen AI and ML, and are in the process of developing improvements or new uses



Power and utilities use cases

Power and utilities encompass generation, transmission, and distribution, each with distinct priorities. Generation focuses on load optimization, plant efficiency, and demand forecasting. Transmission aims to reduce line losses and track assets, while distribution includes grid balancing, smart metering, predictive maintenance, and managing distributed energy and customer service.

Power and utilities companies are looking to build physically accurate grid digital twins to optimize power generation, transmission, and distribution for reliability and resiliency. These organizations need the ability to forecast renewable energy generation from wind and solar with high-resolution, kilometer scale climate and weather simulations.

Accelerated computing enables optimized renewable energy generation and prediction of environmental impacts using AI models, trained on high-resolution weather and sensor data. Power plant operations are optimized using gen AI through real-time data analysis and predictive maintenance leveraging digital twins, ML, deep learning, and natural language processing. Accelerated computing enhances the speed of these predictions, enabling faster decision-making, leading to improved grid stability and a reduction in downtime.

In the transmission stage, HPC and AI are being used to detect anomalies, predict line overloads, and identify potential failures using AI models, power flow simulations, and stability assessments across grid networks.

Gen AI improves transmission by predicting failures and optimizing power flow using real-time data and advanced simulations to support grid operator decisions during both routine operations and emergencies. Simulation times are reduced from hours to minutes using accelerated computing, enabling operators to respond quickly and prevent power blackouts.

HPC and AI accelerate demand forecasting, fault detection, network simulations, and distributed energy workloads for consistent, uninterrupted power distribution. Gen AI is enhancing power distribution by enabling real-time decision-making through advanced grid digital twins, deep learning models, and weather simulation models to improve grid stability and boost operational efficiency. Accelerated computing allows AI applications to deliver real-time fault localization alerts and service restoration data to improve grid reliability and customer satisfaction at the edge.

AWS and NVIDIA for power and utilities

Amazon Web Services (AWS) and NVIDIA enable power and utilities companies to modernize power generation, transmission, and distribution with secure, software-defined infrastructure. AWS provides scalable cloud services that lower operating costs and support advanced analytics, while NVIDIA delivers a full-stack accelerated computing platform for grid-edge intelligence.

Together, AWS and NVIDIA provide power and utilities companies with the tools needed to improve grid reliability, balance electricity demands in real time, integrate renewable energy, and better prepare for climate and weather impacts.



Solutions

Data driven insights

AWS is helping renewable energy companies expedite time-to-value by providing the data foundation for wind, solar, and battery energy storage systems. [Renewables Data Lake & Analytics](#) is a cloud-native solution to monitor and optimize renewable generation fleets at a relatively low cost, at scale, for many millions of tags.

With NVIDIA Omniverse libraries, utilities can build physically accurate digital twins of their power generation facilities to optimize operations with greater intelligence and autonomy.

Real-time performance monitoring

Smart meters can improve customer experience, lower operational costs, and enhance grid reliability. [Meter Data Analytics 2.0](#) unlocks meter data from silos so you can cost-effectively hydrate an Amazon Simple Storage Service (Amazon S3) data lake and use AI, ML, and analytics to uncover valuable insights.

Key tools from NVIDIA, available on AWS, accelerate energy analytics workloads including:

- RAPIDS Accelerator for Apache Spark, which transforms existing pipelines without any code changes.
- NVIDIA cuDSS, a GPU-accelerated direct sparse solver for real-time applications such as grid dispatch simulation.
- NVIDIA cuOpt, a GPU-accelerated solver for decision optimization that can help utilities companies drive significant cost savings.

Transform customer engagement

[Amazon Connect](#), an AI-native contact center, accelerates innovation and transforms every customer interaction. Energy distributors can turn every customer touchpoint into a deeper relationship and better outcome.

Improved production forecasting

AWS and NVIDIA offer tools to rapidly calculate energy production at scale, based on ML algorithms trained with [AWS Batch](#) and AI-driven meteorology enabled by the NVIDIA Earth-2 development platform for climate science. Climate scientists can potentially explore millions of scenarios at a fraction of the cost of traditional weather simulations.

Customer stories

Électricité de France

Électricité de France (EDF), AWS, NVIDIA, and Aneo are working together to iteratively port the code_saturne computational fluid dynamics (CFD) application used by EDF to AWS, where it is accelerated by NVIDIA GPUs. This tool enables safety assessments and lifetime extensions of power plants through high-fidelity modeling of scenarios that are inaccessible to physical experiments. The project aims to enable code_saturne to use NVIDIA software and hardware platforms, while maintaining its modular architecture to facilitate adaptation to NVIDIA GPUs, including access through the AWS Cloud.⁴

Siemens Energy

To boost efficiency for its energy partners, Siemens Energy is utilizing the NVIDIA Dynamo-Triton Server to deliver a path to autonomous power plants, reducing costs along the way. Siemens Energy supports a wide variety of machine learning from images of scenarios taken with onsite cameras and other sensors for data used in analytics. As a result, it requires a highly scalable inference solution—for handling millions of sensors—able to work with multiple frameworks and massive input streams. Siemens Energy selected Dynamo-Triton due to its ability to meet the multi-framework and multi-model requirements. Siemens Energy runs NVIDIA Dynamo-Triton on AWS for scale and multi-tenancy with plans to run at the edge where data cannot be moved out of the powerplants.⁵



⁴"Streamlining GPU Porting for EDF's Fluid Dynamics Simulations with NVIDIA Nsight Profilers," NVIDIA, June 2025

⁵"Siemens Energy Taps NVIDIA Triton Inference Server for Power Plant Inspections, Autonomy," NVIDIA, November 2021

The AWS and NVIDIA advantage

AWS and NVIDIA have collaborated for over 15 years to continually deliver powerful, cost-effective, and flexible accelerated computing solutions for energy industry customers. AWS and NVIDIA deliver industry-leading accelerated computing hardware and software with the scale, stability, and security you require to run AI workloads with confidence. AWS for Energy and Utilities also offers deep industry expertise and extensive energy infrastructure knowledge.

Boost application performance and enable innovation

The combination of AWS's global reach, secure and reliable cloud infrastructure, and breadth of services, accelerated by NVIDIA's full-stack computing platform, provides performance and an end-to-end software stack for AI and accelerated computing in energy.

Transform energy operations

Organizations can solve challenges like predictive maintenance and data analytics at scale, with AWS energy solutions running on high-performance Amazon EC2 instances accelerated by NVIDIA. Or build innovative applications by integrating AWS services for ML, gen AI, and HPC orchestration, together with NVIDIA AI Enterprise software, NVIDIA NIM microservices, and libraries.

Accelerate the clean energy transition

Unlock the innovation and efficiency needed to enable energy initiatives, with cloud-to-edge gen AI applications developed by innovative AI startups, easily accessible through the NVIDIA AI Enterprise platform available in AWS Marketplace.



How to get started

AWS and NVIDIA combine decades of cloud and AI expertise to guide your success from concept to production. Leading energy companies can address many of their key challenges from AI and accelerated computing with AWS and NVIDIA to reduce operating costs, improve profitability, enhance worker safety, and lower environmental impact.

AWS and NVIDIA give you the freedom to choose how you build and deploy AI—with optimized solutions at every step of your journey. Power and utilities companies can harness the combined power of accelerated computing and AI to build a software-defined infrastructure across power generation, transmission, and distribution for enhanced grid reliability and resiliency.

[Learn more →](#)

[Contact an industry expert →](#)

