Resiliency will be re-examined, but few will compromise

Andy Lawrence   20 Feb 2026

OPINION

For decades, the data center industry and its enterprise customers have agreed on the core principles of physical infrastructure resiliency: while it is not possible to avoid failures altogether, the risk and impact of an outage can be minimized through well-established design and operational practices. Notably, these include independence from the power grid and other utilities, and redundancy in critical systems to achieve concurrent maintainability of the facility (i.e., zero scheduled downtime). Further design and operational measures aim to increase facility tolerance against failures.

The industry shares this resiliency philosophy because it has proven effective. According to Uptime Institute survey data (see _Annual outage analysis 2025_), outage rates per site — and certainly relative to IT load — have been trending downwards over time (see **Figure 1**).

Figure 1 Gradual improvements in outage frequency continue

Has your organization had an impactful outage in the past three years ("Yes" respondents)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025
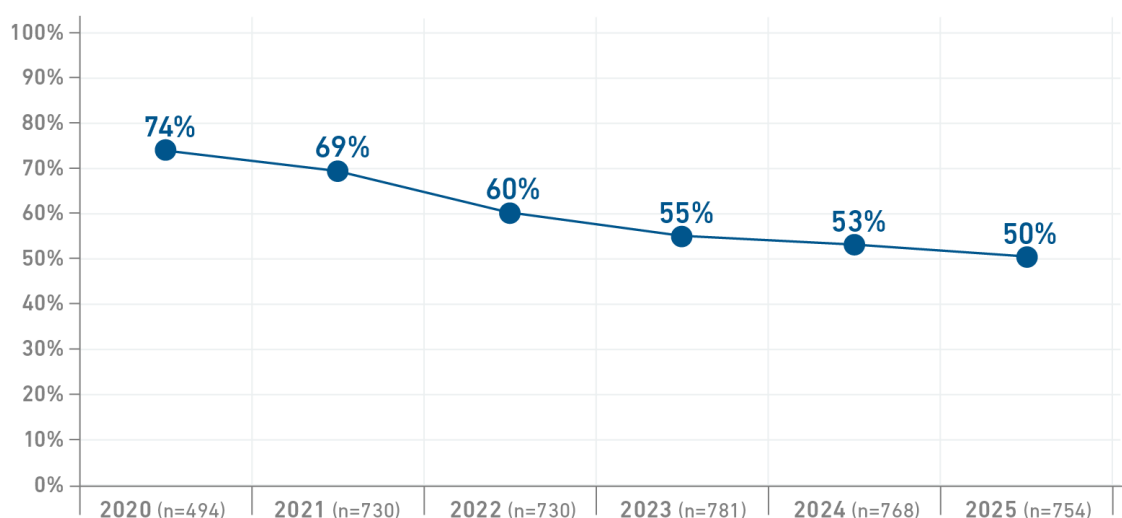
Yet, over the coming years, Uptime Intelligence expects resiliency to re-emerge as a major area of focus and concern for the industry — indeed, the first whispers of this could be heard in various discussions and proposals in 2025. The reasons for this renewed attention on the issue include:

- Structural, long-term challenges with many power grids, coupled with the concentration and scale of data centers within those regions, are increasing the likelihood of grid instability and failures, some of which will be major and sustained. Increased regulatory intervention is also likely.

- The rise of AI is changing the load profile, behavior and distribution of IT, making strategies such as availability zones (AZs) and distributed resiliency more difficult or more expensive to achieve.

- Growing reliance on expensive, large-scale on-site power generation will force some operators to share their generating capacity. This will require closer collaboration with transmission service operators (TSOs) to manage risks and capacities more actively, and to balance operators' resiliency goals with the needs of their grid partners.

- The increasing use of alternatives to diesel generators — such as fuel cells, gas turbines and large-scale batteries — coupled with a more statistical and dynamic approach to power capacity and redundancy, is encouraging some designers to consider approaches that may offer reduced resiliency but can be delivered relatively quickly and at lower costs (Google data center design spin-off Verrus is an example; its designs focus on software and batteries rather than generators).

All this is happening against a backdrop of increasing jeopardy. While large outages have become less likely relative to the size of the industry, the consequences of a failure have risen — a trend also reflected in Uptime Intelligence data (see *Uptime Institute Global Data Center Survey 2025*). The ripple effects of a major site outage can extend far and wide: the loss of a core cloud or software service, for example, can affect thousands of third-party applications and services, while the automated rerouting of traffic may result in further cascading service failures due to network congestion.

Regulators have noticed these risks to national economies. Some have designated data centers as critical national infrastructure as an initial step toward more intervention. Others are planning new resiliency-focused legislation.

Concentration and scale

The regional concentration of IT service providers and fiber network operators was once a major concern for the resiliency of the internet, but this has now largely passed. However, high concentration has become a risk for data center regions themselves, such as Northern Virginia (US) or the Dublin (Ireland) area.

There are three specific concerns. First, growing electrification, aging transmission equipment,

and an increasing share of intermittent generation in the energy mix are each creating conditions for grid instability. In September 2025, the US Department of Energy (DoE) warned that grid power outages could increase by a factor of 100 by 2030.

Most data centers are designed to handle power outages by switching to engine generators. However, more frequent grid disturbances and outages will place greater stress on critical power systems. A major grid failure can also lead to refueling delays, increase particulate emissions in populated areas, and operational problems caused by loss of communications and staff shortages.

Second, data centers have become part of the problem. Large campuses and concentrations mean that data centers are stretching grid supply, and their behavior can affect grid stability (see *Power companies act to stop data center-induced blackouts*). Two examples serve as a warning: voltage fluctuations in Northern Virginia in July 2024 and in Ireland in December 2022 caused many data centers to disconnect from the grid near simultaneously, causing power surges (oversupply) that almost led to blackouts in both cases.

Heeding these lessons, authorities and TSOs in several countries — including US, Australia, Singapore, the UK, Ireland and much of Europe — are now working on new grid connection rules (see *Ireland's new grid rules signal shift in data center roles*). These will oblige operators to tolerate higher voltage fluctuations and to avoid, or slow down, load disconnection during grid disturbances.

The third concern is that major grid outages may strain, or even cripple, fuel supplies. During the Spanish blackout in early 2025, Airbus's Madrid data center experienced a critical diesel shortage as fuel suppliers became overwhelmed. A major power outage affecting London Heathrow in the UK also caused some operators to review their access to fuel. Storing larger on-site fuel reserves is usually the first option, but operators are increasingly focusing on their arrangements for emergency resupplies. In some jurisdictions, national critical infrastructure classification will help secure priority access to fuel during large-scale outages.

IT resiliency: cost calls for more dynamism

AI is also creating challenges for those whose resiliency strategy relies on partly spreading work and risk across multiple data centers.

For decades, the practice of high-availability and fault-tolerant clustering across multiple data centers has been used for select mission-critical applications. The emergence of cloud AZs, together with affordable IT hardware and novel application development styles, has popularized the idea of multi-site resiliency for a much broader range of applications. Increasingly, new IT applications are operated in a scale-out fashion, distributed across several facilities using load balancers.

However, there are challenges. These include software design complexity, the difficulty of mapping and eliminating single points of failure, and the cost of wide-area networking required for

high-speed data replication (see *Cloud availability comes at a price*).

Distributed resiliency techniques are neither ubiquitous nor bulletproof, but they have proven successful for services designed for the cloud era. IT and distributed resiliency now increasingly complement site-based resiliency and, as these techniques become more robust, they may replace it in a few use cases.

Unfortunately, though, AI may have made this more difficult. For AI training, distributed resiliency is a non-starter — at least with current model designs — due to the need for very low latency across processing nodes. For inferencing, this is less clear; however, the high cost of IT for inferencing and the growing need for customized hardware configurations will likely make distributed resiliency equally challenging.

Depending on their risk tolerance and IT architectures, some organizations will continue to absorb the facility costs of high physical infrastructure resiliency across all applications, including AI compute. Others, however, will feel encouraged to examine more thoroughly what level of resiliency each application or service really needs, starting with larger AI workloads. As analytical tools and data improve, making these distinctions — and creating multiple, virtual or dynamic "tiers" — may become more practical than in the past.

Large dense sites will pay for resiliency

Since 2021, Uptime Intelligence has identified development plans for more than 300 data center campuses worldwide exceeding 100 MW — some even surpass 1 GW. Hundreds more are in development at a smaller scale, although these are still very large by historical standards.

Providing full-sized emergency backup power at this scale entails enormous cost, often amounting to tens of millions of dollars to install. However, capital costs are only part of the problem. The process of sourcing, testing and maintaining engines and fuel systems is increasingly laborious. Then there is the logistical issue of resupplying fuel at such a large scale.

In the battle between OpenAI, Google, Meta and others to train the largest and best-performing generative AI models, resiliency has not been a high priority — "moving fast and breaking things" is the Silicon Valley ethos. This has promoted a perception that AI training facilities running batch loads can be built without concurrent maintainability or even emergency backup power.

However, this risk-taking behavior is unlikely to be followed by the majority of new data center builders, despite the huge expense and challenges of higher levels of resiliency. Uptime Institute is aware of, and in some cases involved in, the construction of very large facilities, built for AI, that do not compromise on resiliency.

One clear reason for this is that some of the larger facilities will build their own (or collocate with) prime power generation and distribution infrastructure. Adding redundancy to prime generation (in N+1 configurations) will likely cost less, relatively, than providing mostly unused standalone backup power for the grid, particularly since some of this capacity can be sold to the grid as part

of demand response schemes. In any case, IT infrastructure costs so comprehensively dwarf the cost of power provision and redundancy that any outage carries a heavy cost in terms of reduced utilization, revenues, service levels and customer relationships. Indeed, lessons have been learned from the cryptomining sector, where prolonged outages at data centers without redundant power proved costly — such as in early 2024 at Applied Digital (now an AI data center provider), which resulted in more than $60 million in lost production.

This is underlined by a second, broad-based observation. Except for a small number of very high-density, HPC-like facilities, most data centers will likely support, or need the capability of supporting, a rich mix of workloads. Because technical requirements several years into the future are highly uncertain, operators will want to ensure that facility investments remain durable. Most data center customers, including hyperscale and other IT service providers, will continue to require strong resiliency guarantees.

ABOUT THE AUTHOR

Andy Lawrence

20 Feb 2026

Andy is a founding member and the Executive Director of Research for Uptime Institute Intelligence, which analyzes and explains trends shaping the critical infrastructure industry. He has extensive experience analyzing developments in IT, emerging technologies, data centers and infrastructure, and advising companies on technical and business strategies.

alawrence@uptimeinstitute.com

About Uptime Institute

Uptime Institute is the Global Digital Infrastructure Authority. With over 4,000 awards issued in over 122 countries around the globe, and over 1,100 currently active projects in 80+ countries, Uptime has helped tens of thousands of companies optimize critical IT assets while managing costs, resources, and efficiency. For over 30 years, the company has established industry-leading benchmarks for data center performance, resilience, sustainability, and efficiency, which provide customers assurance that their digital infrastructure can perform across a wide array of operating conditions at a level consistent with their individual business needs. Uptime's Tier Standard is the IT industry's most trusted and adopted global standard for the design, construction, and operation of data centers.

Offerings include the organization's Tier Standard and Certifications, Management & Operations reviews and assessments including SCIRA-FSI financial sector risk assessment, the Sustainability Assessment, and a broad range of additional risk management, performance, availability, and related offerings. Uptime Education training programs have been successfully completed by over 100,000 data center professionals, such as the much-valued ATD (Accredited Tier Designer) and AOS (Accredited Operations Specialist). The Uptime Education curriculum has been expanded by the acquisition of CNet Training Ltd. In 2023.

Uptime Institute is headquartered in New York, NY, with offices in London, Sao Paulo, Dubai, Riyadh, and Singapore, and full-time Uptime professionals based in over thirty-four countries around the world.

For more information, visit www.uptimeinstitute.com