parametrix

Cloud Outage Risk Report 2025

Executive Summary

The rapid expansion of **AI workloads** is driving unprecedented demand for compute capacity, **increasing infrastructure complexity and amplifying exposure to performance and availability risks.**

The number of ‘Critical’ cloud service interruption events among AWS, Azure, and GCP **declined by 8.2%, from 49 in 2024 to 45 in 2025.**

Although the number of critical events declined slightly, outages in key regions such as AWS us-east-1 and GCP us-central1 led to disproportionately large disruptions.

Outages affecting **Compute services exceeded 100 hours in 2025,** consistently representing the largest share of total downtime duration since 2022.

Emerging service categories, particularly AI and analytics, experienced elevated downtime during their rapid development and scaling phase, **highlighting new areas of risk within the cloud stack.**

North America accounted for more than half of all events both in 2024 and in 2025, a significant rise from 39.9% in 2023.

The first half of 2025 digital supply chain disruptions were relatively low and consistent, however **the second half of the year shows a series of high-impact events that affected the entire digital ecosystem.**

Expanding cloud infrastructure in the AI era

Cloud computing is now foundational for the global economy. Its importance intensified in 2025 as demand was accelerated by the rapid adoption of Artificial Intelligence (AI).

AI workloads, particularly generative AI and machine learning models, are dramatically increasing the need for compute power, storage, and low-latency connectivity. To meet this demand, **cloud service providers are investing at unprecedented levels, building new data centers, expanding existing campuses, and pushing infrastructure to operate at greater scale and density than ever before.**

While this growth fuels innovation, it also introduces new forms of risk. Despite its ethereal name, the cloud is a vast, interconnected physical system dependent on power availability, cooling functions, human operations, software integrity, and geographic concentration.

As the complexity of this physical system multiplies, the potential for disruption increases. The impact of cloud outages cascades across digital supply chains to disrupt thousands of businesses simultaneously, and generate material financial losses which may span sectors and territories.

At Parametrix, we believe that understanding this evolving risk landscape starts with data.

The Parametrix Cloud Monitoring System conducted 16.5 billion tests across 500 data centers operated by AWS, Azure, and GCP worldwide in 2025. It delivers insights into where, how, and why outages occur. That data enables enterprises, insurers, and risk managers to move beyond assumptions and toward evidence-based decision-making.

The 2025 Cloud Outage Risk Report reflects this mission. As reliance on cloud services continues to deepen, resilience can no longer be an afterthought. It must be measured, managed, and insured with the same rigor applied to other forms of critical infrastructure risk. **Parametrix remains committed to providing the transparency, data, and solutions needed to support that goal.**

> **Cloud service providers are investing at unprecedented levels - pushing infrastructure to operate at greater scale and density than ever before.**

Cloud outages in 2025

After a significant escalation in cloud downtime from 2022 through 2024, Parametrix detected a notable decrease in 2025. The improved annual reliability is explained primarily by a quiet first half, when no significant events occurred.

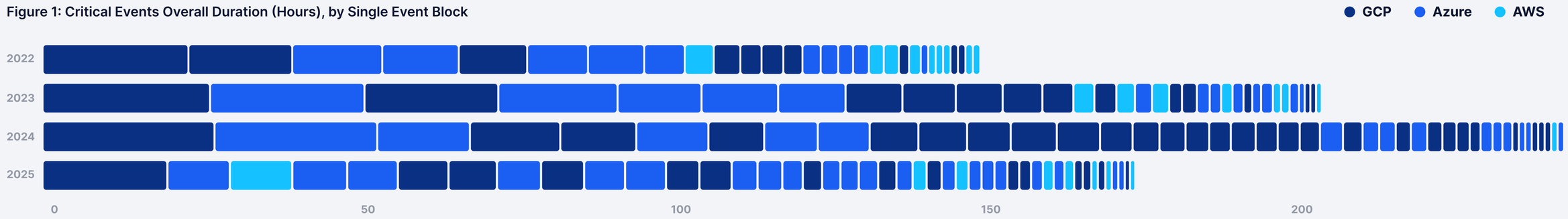
However, major outages began to happen around mid year. **Further, despite recording fewer downtime hours than in previous years, the impact of outages on individual businesses was greater in 2025.** This is because sustained interruptions occurred in some of the largest and most critical regions, including AWS us-east-1 and GCP us-central1.

The aggregate duration of “critical” downtime events - those which caused a complete shutdown of services or a significant service interruption, and therefore had a widespread impact on users - rose steadily from 150.9 hours in 2022 to a peak of 244.8 hours in 2024. However, 2025 marked a turning point in this multi-year surge in high-impact outages. **Total downtime decreased by approximately 28% to 175.3 hours** (Figure 1).

The number of critical interruptions also decreased, from 49 in 2024 to 45 in 2025. The drop of 8.2% returned this metric to the level recorded in 2023.

The volatility was driven largely by provider-specific performance. For instance, GCP experienced a difficult 2024, with 148.4 hours of downtime, but improved significantly in 2025. In contrast, after a highly stable 2024 AWS suffered a sharp increase in outage duration in 2025, including the widely disruptive us-east-1 event on October 20, 2025.

Figure 1: Critical Events Overall Duration (Hours), by Single Event Block



Breakdown by cloud service

Parametrix monitors the functionality of CSPs' various services. The landscape of service reliability has shifted, though core infrastructure remains a primary driver of downtime.

Compute services have consistently accounted for the most significant portion of outage duration, exceeding 100 hours annually every year from 2022 through 2025, with a peak of 136.5 hours in 2023. **Database** services also remained a steady contributor to critical interruptions, peaking at 76.9 hours in 2023 and 66.4 hours in 2024, before showing improved stability by dropping to 43.2 hours in 2025

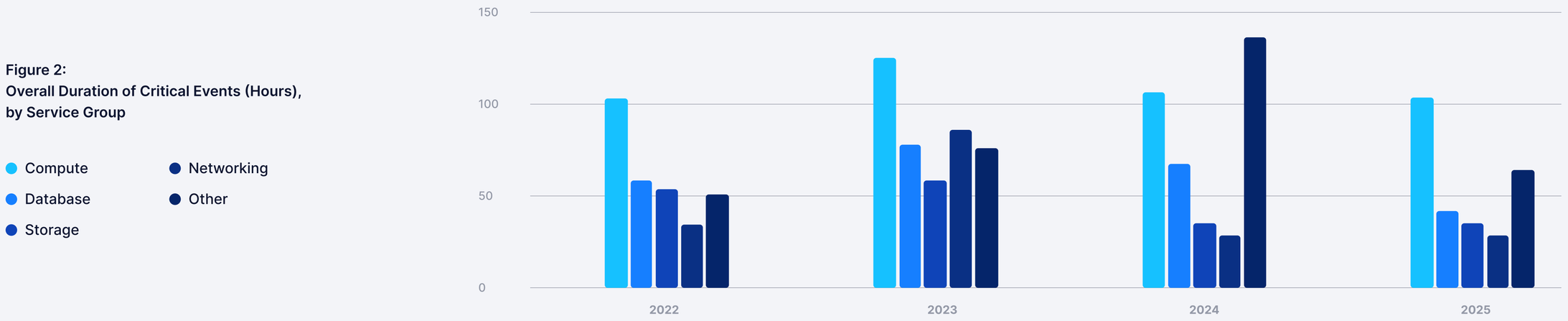
Networking and Storage services demonstrated notable recoveries. After a severe networking spike to 83.5 hours in 2023, downtime dropped sharply to 33.7 hours in 2024 and further stabilized at 28.9 hours in 2025.

Storage followed a similar positive trajectory, maintaining lower outage durations and falling from 62.3 hours in 2023 to 37.5 hours by 2025.

Notably, the broader **Others** category saw an increase in hours of downtime from 2022 to 2024, before going down in 2025.

When this category is separated, it reveals that **this upward trend was heavily driven by AI and Analytics services, reflecting the massive development, scaling, and integration of these technologies during 2023 and 2024.** AI services, which initially caused minimal downtime in 2022, jumped significantly and maintained consistently high downtime levels throughout the development boom in the subsequent years.

Figure 2:
Overall Duration of Critical Events (Hours), by Service Group



Breakdown by geography

The number of downtime events experienced varies significantly by region. **North America accounted for more than half of all events both in 2024 and in 2025, a significant rise from 39.9% in 2023.** This dominance is unsurprising, since North America is the largest geography for cloud usage. It hosts the highest density of live websites and data center infrastructure.

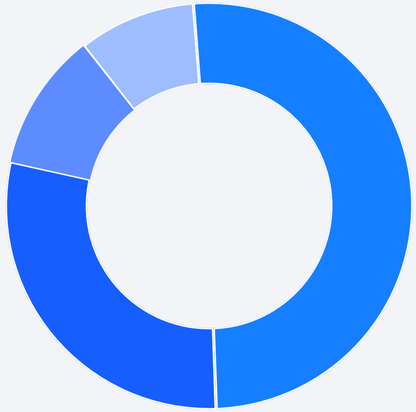
Europe followed a more volatile trajectory. After a temporary reprieve in 2024, when its share of incident frequency dropped, the continent reverted to 2023 levels of instability in 2025.

In contrast, **Asia** has demonstrated a clear and positive long-term trend. The frequency of interruptions there has steadily declined year-over-year from 2023 through 2025.

The **Rest of World** remained largely static, maintaining a consistent, low baseline of activity throughout the three-year period.

> **North America accounted for more than half of all events both in 2024 and in 2025.**

Figure 3:
Geographical Split,
by Duration

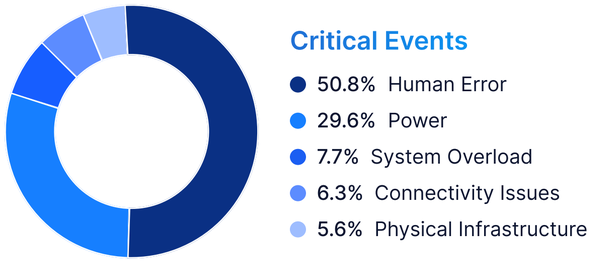


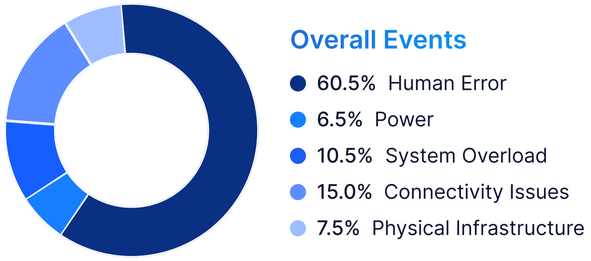
2025

- 50.8% North America
- 29.0% Europe
- 10.9% Asia
- 9.3% Rest of World

2024	2023	2022
61.5% North America	56.7% North America	43.5% North America
19.9% Europe	13.2% Europe	30.1% Europe
11.2% Asia	16.8 Asia	10.6% Asia
7.4% Rest of World	13.4% Rest of World	15.8% Rest of World

Figure 4:
Number of events by Root Cause (2022-2025)





Breakdown by root cause

The distribution of root causes shifts significantly when comparing all cloud events to those classified as “Critical” (Figure 4). While Human Error remains the leading cause in both categories, Power outages exhibit a disproportionately high impact on critical event frequency compared to their overall occurrence.

Human error accounts for ~61% of all recorded events. While still the leading cause, the share of human error drops to 51% when only critical events are considered. This suggests that while human mistakes are frequent, they are slightly less likely to result in critical-level failures compared to other causes like power loss.

Power, the critical differentiator: the most significant disparity in the distribution appears in power-related incidents. These issues are relatively rare in the general dataset, where they are responsible for only 6.5% of the total event set. However, **power incidents account for ~30% of critical outage events.**

When power incidents do occur, they have a high propensity to trigger critical performance interruptions or total service failures.

Connectivity Issues account for 15% of all events, but only 6% of critical events.

System Overload represents 11% of all events, but drops to 8% of critical events.

Physical Infrastructure, hardware and buildings issues, remain relatively consistent across both categories, accounting for 8% of all events and 6% of critical events.

Parametrix Radar

Parametrix has collected and analyzed more than 54 billion cloud availability data points that defines the historical performance of cloud services. We also monitor directly the service availability of more than 7,000 top SaaS, PaaS and IaaS providers. Parametrix backs up this unique hard data with extensive expertise in system failures and business interruption losses to identify cloud outages, determine their extent, and estimate their impact (Figure 5).

A distinct improvement in overall cloud stability was realized in 2025. Figure 5 corroborates this trend, displaying a "silent" first half of the year, when impact levels remained relatively low and consistent. However, **the second half of the year shows a series of high-impact events that affected the entire digital ecosystem.**

In 2025, every major cloud provider experienced at least one impactful event, along with significant disruptions in critical digital supply chain services:

- **GCP (June 12):** The first major spike occurred mid-year, triggered by a faulty quota policy update causing widespread API failures across core services like Compute Engine.
- **Cloudflare (August 21, November 18, December 5):** Distinct spikes linked to Cloudflare outages highlight the critical role of Content Delivery Networks (CDNs)- when they fail, access to applications can degrade or stop entirely.
- **AWS (October 20):** The largest spike in impacted companies (and the biggest event recorded by Parametrix Radar) occurred during the AWS us-east-1 outage, caused by failures in DynamoDB's DNS automation, severed connectivity for a number of dependent services.
- **Azure (October 29):** Shortly after the AWS incident, Azure experienced a global outage involving Azure Front Door caused by a configuration change. It contributed further to the volatility of the final quarter of the year.

This sequence of events illustrates that while the total hours of downtime decreased, the impact of specific, isolated incidents remained high, with no major player immune to critical failure in 2025.

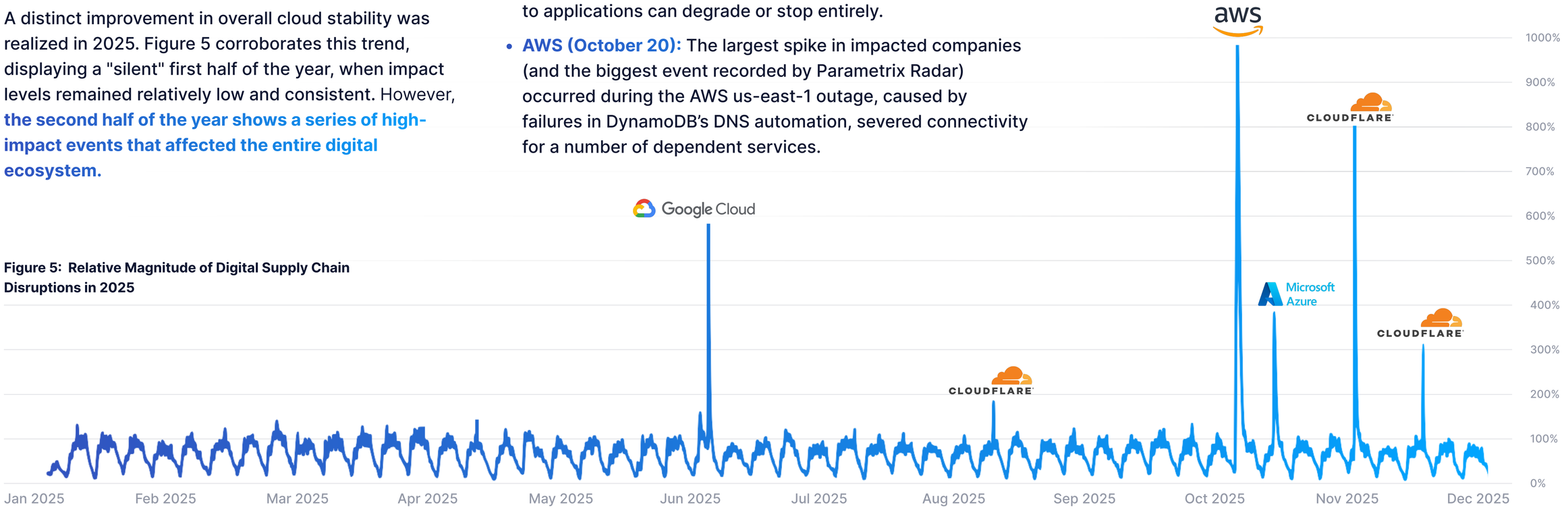


Figure 5: Relative Magnitude of Digital Supply Chain Disruptions in 2025

Impacts on the cyber insurance market

Today almost all companies have some degree of cyber exposure. Even organizations with limited internal technology operations rely heavily on third-party digital services for payments, communications, data storage, customer engagement, and core business processes. As a result, cyber risk has evolved from a niche concern into a universal business risk.

One of the most significant shifts in recent years has been the growing importance of business interruption losses stemming from digital system failures. Cloud outages, SaaS disruptions, and other third-party failures within the digital supply chain can halt operations entirely, even when a company's own systems remain intact. These events expose a critical gap between traditional cyber coverage and real-world loss scenarios.

As reliance on third-party services continues to rise, so does the frequency and severity of contingent business interruption events. A single outage can impact thousands of insureds simultaneously, creating aggregation risk for insurers and reinsurers.

An additional layer of exposure is embedded within the SLAs that govern cloud, SaaS, and data center services.

These contracts define uptime commitments and performance thresholds which, when breached, can trigger financial penalties, service credits, or downstream contractual liabilities. **In practice, SLA breaches often coincide with cloud outages and digital system interruptions, amplifying business interruption losses beyond the immediate operational impact.**

This has important implications for the cyber insurance market. Underwriters are increasingly challenged to quantify exposure tied to cloud providers, data center regions, and shared digital dependencies. Claims teams face complexity in determining triggers, attribution, and loss measurement when outages stem from third-party infrastructure failures rather than malicious attacks. At the same time, insureds are seeking more certainty.

Businesses want coverage that responds predictably to downtime events, without lengthy disputes over causation or policy interpretation. This has driven interest in new insurance structures that better align with how digital interruptions actually occur.

Parametrix Cloud Stability Index

The Parametrix Cloud Stability Index **(PCSI), provides a benchmark of CSP performance for the year.** Based on the proprietary Parametrix Cloud Modeling System (PCMS), the Index assesses factors including the number of service interruption events per region, their relative severity, duration, and the broader impact of each incident.

The PCSI classifies stability into five grades, from “A” (scores of 0 to 20, representing the most stable regions) to “E” (scores of 80 to 100, reflecting the least stable regions). These scores represent relative performance during the measured period and are not projections of future risk or reflective of past years’ performance. The calculations power Parametrix’s risk modeling, product innovation, and cloud resilience consulting for clients and partners, supporting our mission to quantify and manage cloud risks.

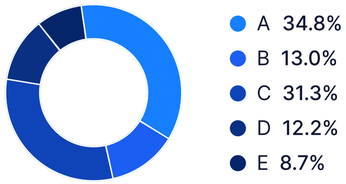


Figure 6: Market Leaders Percentage of Regions by Parametrix Cloud Stability Index Grades, 2022-2025

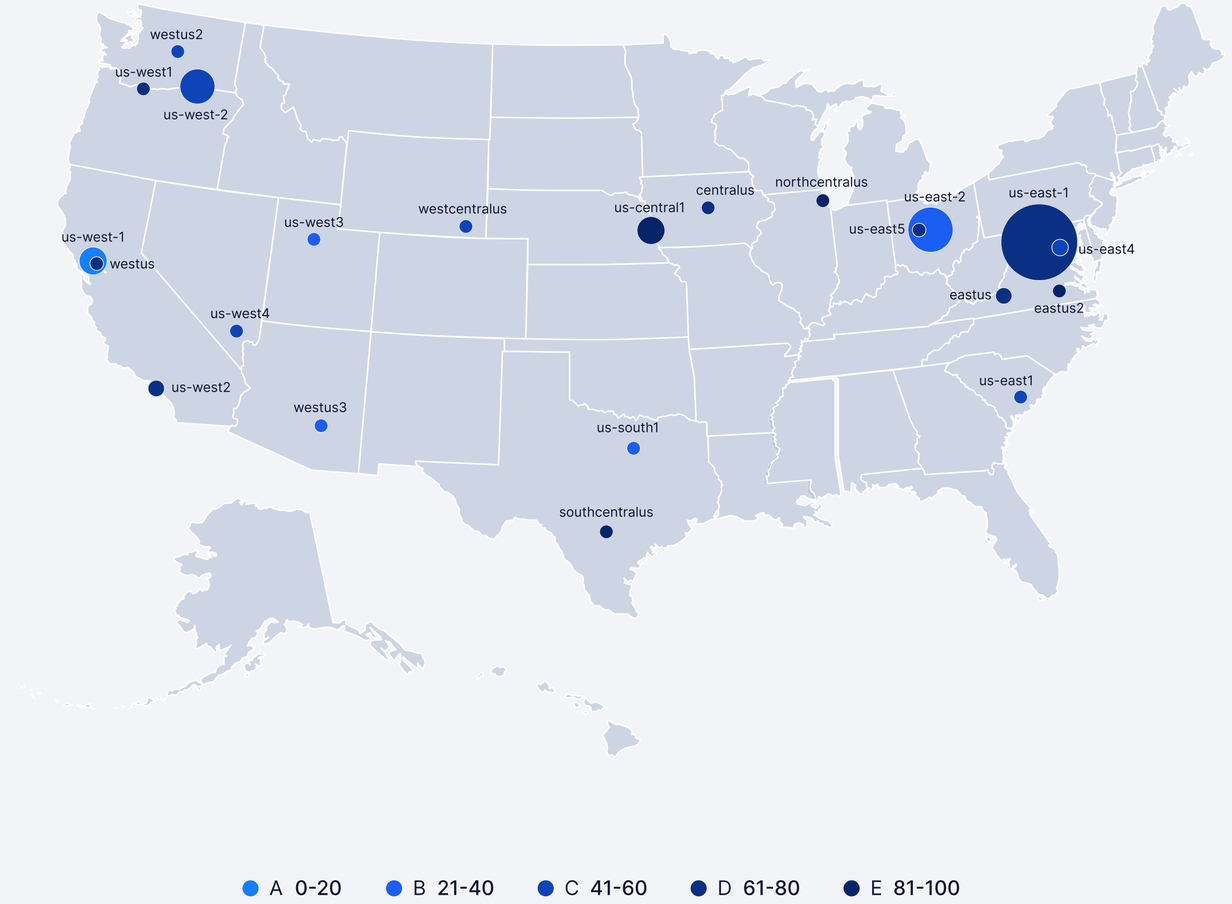


Figure 7: US Region Usage Level (Size) and Parametrix Stability Index (Color)

Cloud outage event analysis

Google Cloud

Start Date
12 June 2025

Duration
2 hours, 45 minutes

Cloud Region(s)
us-central1 (Council Bluffs , IA)

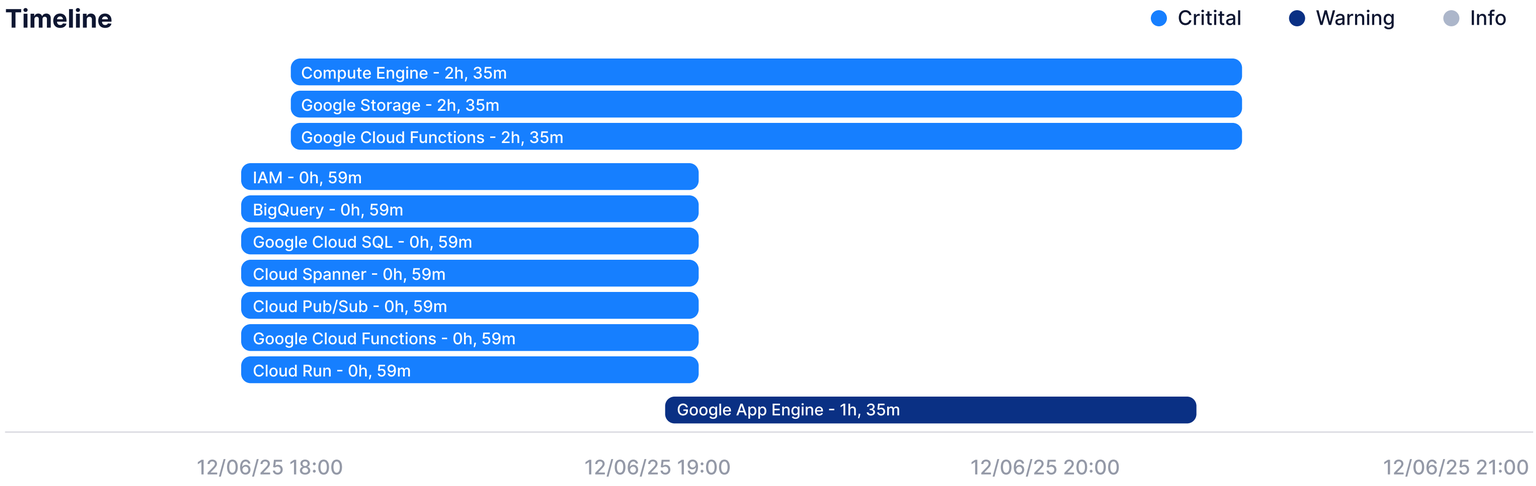
Root Cause
Faulty quota policy update in Service Control

Impacted Services Include:
IAM
Google Cloud Storage
Compute Engine
BigQuery
Google Cloud SQL
Cloud Spanner
Cloud Pub/Sub
Google Cloud Functions
Cloud Run

On June 12, 2025, Google Cloud Platform (GCP), along with Google Workspace and Google Security Operations, experienced a major outage that impacted customers around the world. The incident lasted just under three hours and disrupted access to some of the most widely used cloud services, including Gmail, Google Cloud Storage, BigQuery, and more. The most severely impacted region was us-central1, where API failure rates for core services approached 100% and persisted for 2 hours and 35 minutes.

The root cause of the outage was traced to a faulty quota policy update in Service Control, a core component of Google's API management infrastructure responsible for enforcing usage quotas and managing access to thousands of Google APIs.

Timeline



Amazon Web Services

Start Date
20 October 2025

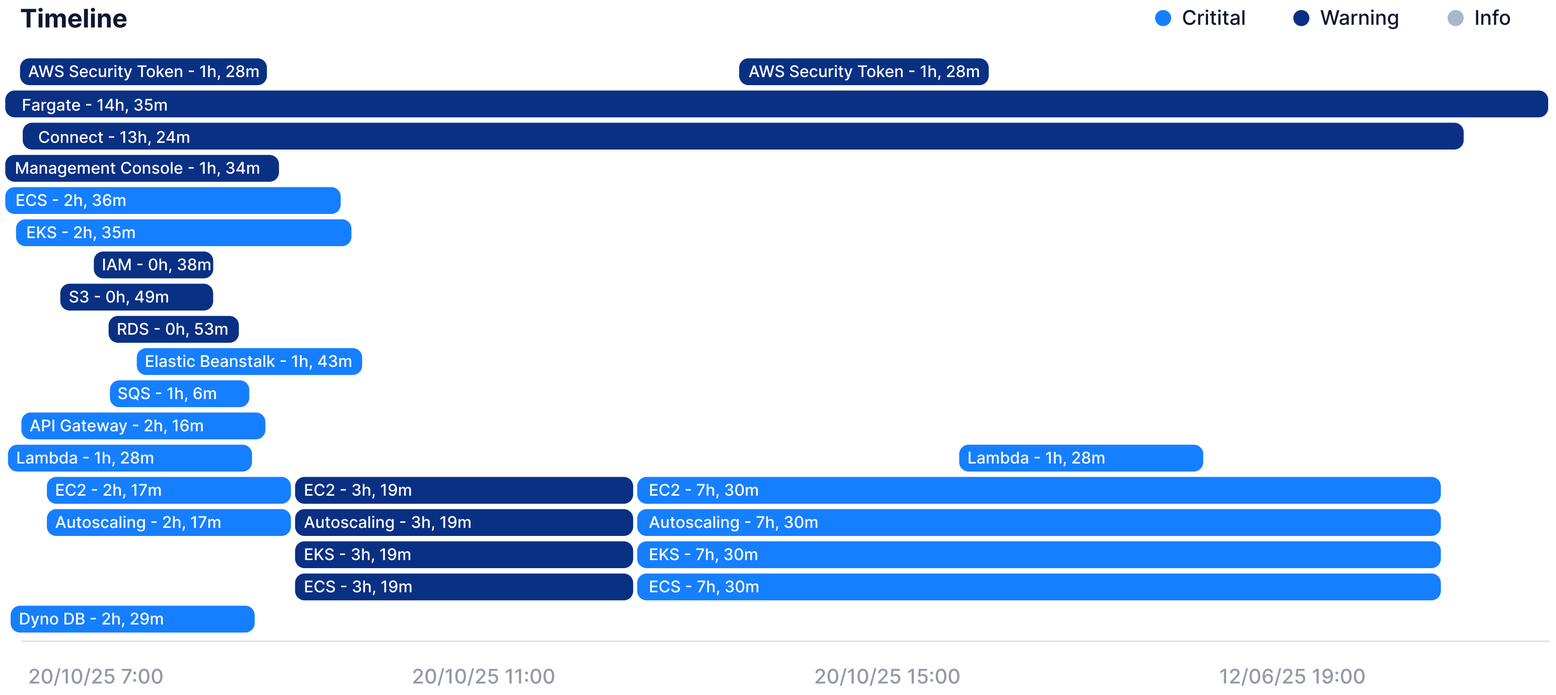
Duration
14 hours, 35 minutes

Cloud Region(s)
us-east-1 (N. Virginia)

Root Cause
DNS, automation failure; Issue with subsystem

Impacted Services Include:
Dyno DB
EC2
Lambda
Auto Scaling
EKS

On October 20, 2025, Amazon Web Services (AWS) experienced a major service disruption in its us-east-1 (Northern Virginia, USA) region, one of the largest and most critical cloud regions globally. Parametrix estimates that the financial losses for US companies resulting from the event is between $500 million and $650 million.
The issue originated from failures in its internal DNS resolution layer, which caused widespread service name resolution errors. In simpler terms, AWS's internal system for translating domain names into IP addresses stopped responding, breaking communication between key infrastructure components.



Microsoft Azure

Start Date
29 October 2025

Duration
8 hours, 23 minutes

Cloud Region(s)
Global

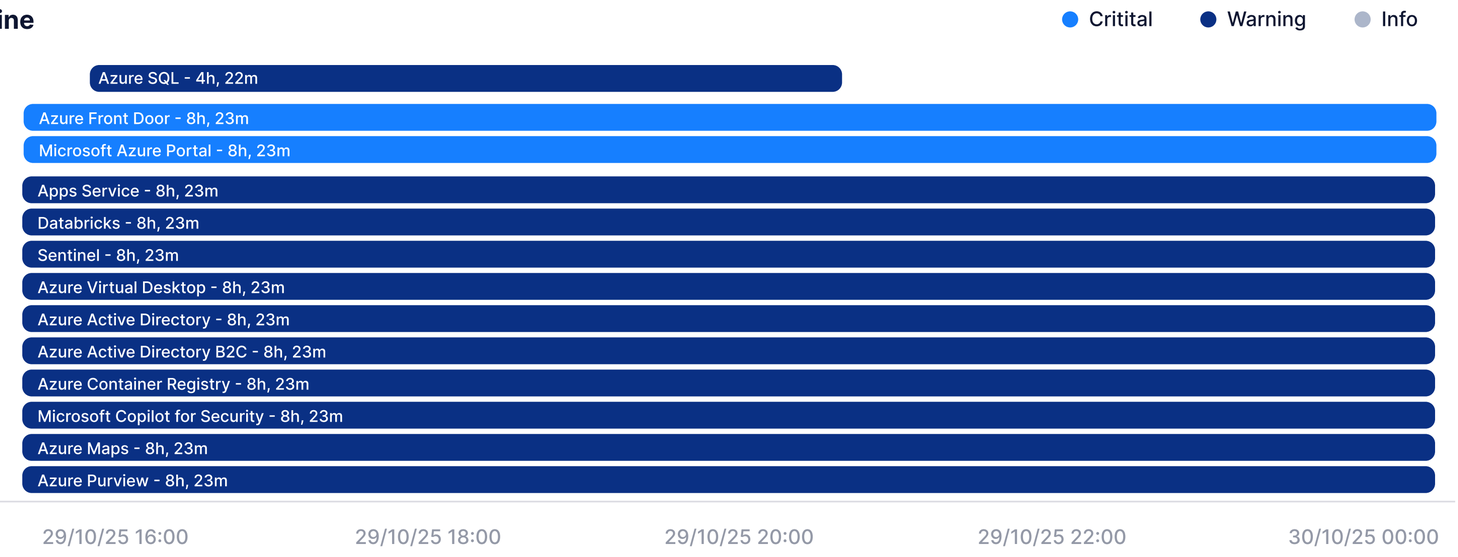
Root Cause
Configuration change in AFD

User Impact
Azure Front Door
Microsoft Azure Portal
Databricks
Azure SQL
Azure Virtual Desktop
Azure Container Registry
Microsoft Copilot for Security

On October 29, 2025, Microsoft Azure experienced a significant disruption stemming from an accidental configuration change in Azure Front Door (AFD), Microsoft's Content Delivery Network (CDN) service. The event lasted from 15:45 UTC on October 29 until 00:05 UTC on October 30, affecting users and numerous services worldwide.

Because AFD operates at the internet edge, the misconfiguration led to widespread latency, timeouts, and connection errors across public web endpoints. The primary effect was a reduction in the availability of websites and web applications, causing slow load times, intermittent 4xx/5xx responses, and portal sign-in issues for end users.

Timeline



parametrix

Parametrix is the leading provider of digital system interruption insurance solutions.
The company specializes in underwriting parametric insurance that protects against the financial cost of technology and digital infrastructure downtime. Leveraging a proprietary network of monitoring systems, we collect and analyze real-time, granular data on the performance and availability of critical infrastructure — including data centers, SaaS providers, and cloud services — to accurately assess risk and deliver fast, transparent claims payments. Our solutions enable businesses, data center stakeholders, and (re)insurers to quantify, manage, and transfer the financial risks of downtime with unmatched precision. Parametrix is a Managing General Agent and Lloyd's Coverholder, with policies backed by major A-rated global insurers, and is headquartered in New York.

www.parametrixinsurance.com

parametrix-insurance

info@parametrixinsurance.com