FROST & SULLIVAN

What Determines Cloud Resilience?

A Stable Cloud Foundation Underpins Sustained Business Growth

FROST & SULLIVAN WHITEPAPER

CONTENTS

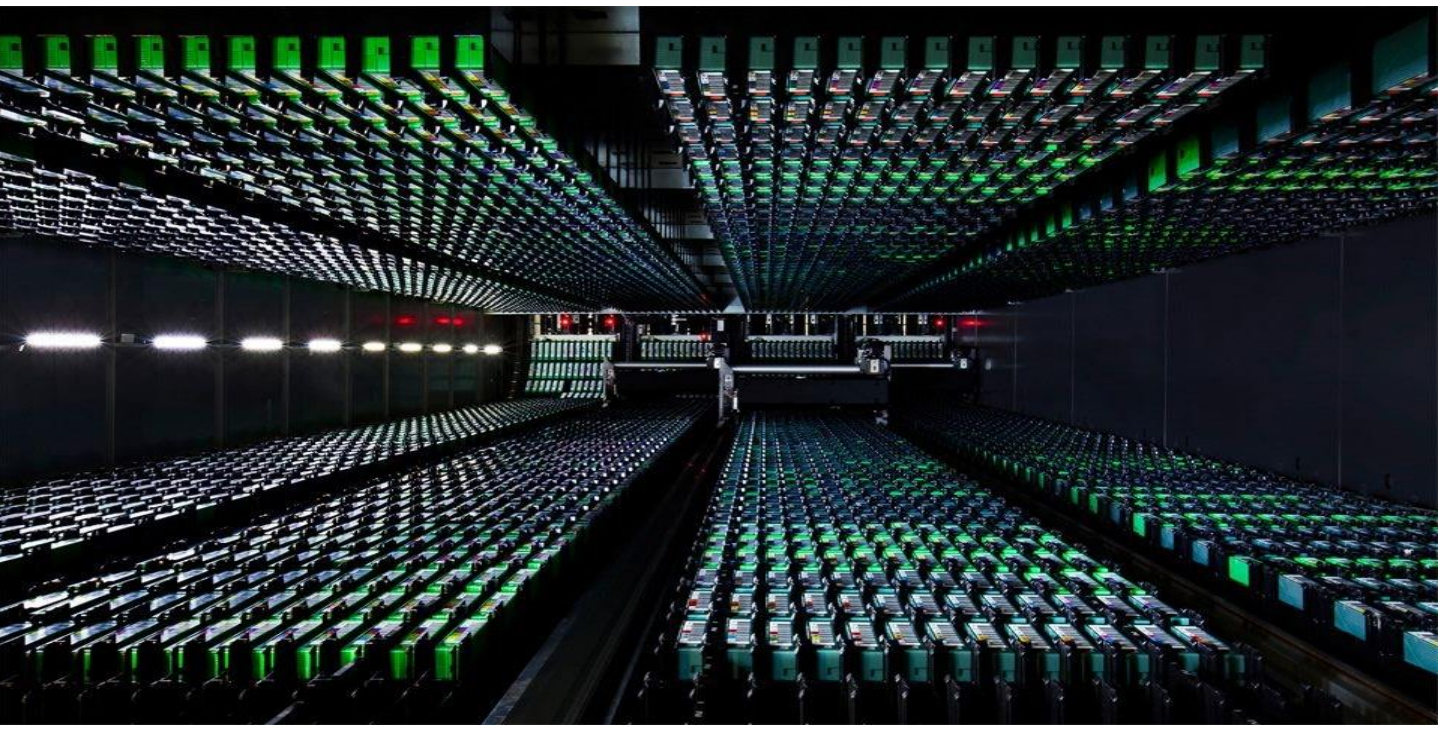
Introduction

Digital infrastructure forms the foundation of all digital operations. Whether organizations rely on on-premises data centers, private clouds, or public clouds, any service disruption or failure can trigger system and software malfunctions, potentially leading to business failure. Such incidents often result in significant economic losses, reputational damage, and customer dissatisfaction. For critical public services, disruptions may even provoke social unrest or other severe consequences.

According to a Frost & Sullivan survey in 2024[1], each minute of digital infrastructure downtime can incur costs ranging from thousands to tens of thousands of RMB. In sectors such as finance and e-commerce, the loss per minute may exceed one hundred thousand RMB. Clearly, ensuring resilience of digital infrastructure is a top priority for organizations across industries.

How can resilience of digital infrastructure be ensured? Frost & Sullivan's research highlights that the robustness of the architecture design and infrastructure deployment, along with resource configuration and redundancy, is essential to resilience.

To further explore how architecture and resource allocation impact service resilience, we conducted a systematic study of the infrastructure architectures applied by leading cloud service providers (CSPs) in China and their resilience performance. Focusing on mainland China, this study presents key data illustrating the varying levels of resilience among CSPs.



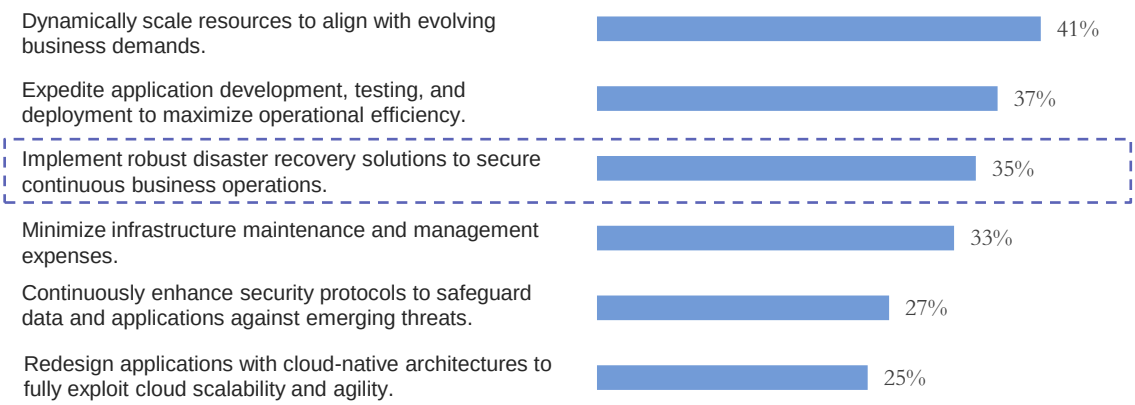
1. In Q4 2024, Frost & Sullivan surveyed about digital business practices across 150 Chinese enterprises spanning a range of industries and company sizes.

This study begins by clarifying key cloud infrastructure terminologies and assessing whether significant differences exist in the terms used by various CSPs. Our analysis focuses on 5 leading providers: Amazon Web Services, Huawei Cloud, Alibaba Cloud, Tencent Cloud, and Microsoft Azure, exploring how each of them designs and deploys cloud infrastructure within China. To evaluate cloud service resilience, we compare critical data on service disruptions and outages reported by these providers from January 2023 to March 2025. Before presenting the comparison, we detail our data sources, evaluation criteria, and research background. We also provide a comparative analysis of resilience between on-premises data centers and cloud services. The study concludes with actionable recommendations for enterprises seeking to enhance resilience of workloads in the cloud.

Why Does Cloud Resilience Matter

According to a Frost & Sullivan survey[2] conducted in Q1 2025 on cloud adoption among Chinese enterprises, over 90% rely exclusively on public cloud for non-critical workloads, failing to fully leverage the cloud's potential. A primary reason for hesitating to migrate core workloads is the concern about cloud service resilience. More than 80% of respondents believe that, despite higher upfront and maintenance costs, on-premises data centers may provide stronger service resilience for critical workloads.

Reasons Cited for Cloud Migration by Chinese Enterprises

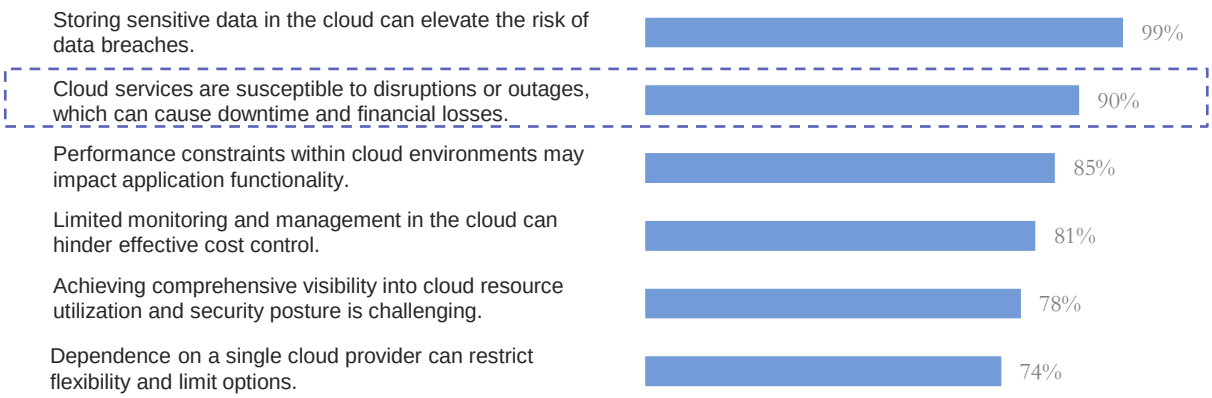


Source: Frost & Sullivan, N=120,
Refer to note 2 for the detailed background of the research.

When migrating business workloads to the cloud, surveyed enterprises prioritize key factors such as elastic scalability of computing resources, modern application development and deployment tools, high availability and disaster recovery solutions, cost-efficient resource, and robust security and compliance controls. Data from the chart above indicate that cloud service resilience and business continuity are prioritized immediately after resource elasticity and development efficiency.

2. In Q1 2025, Frost & Sullivan conducted a detailed, multi-faceted survey on cloud adoption across 120 companies of varying sizes and industries. Respondents were primarily IT procurement decision-makers, operations managers, solution architects, and security engineers.

Concerns among Chinese Enterprises over Cloud Migration



Source: Frost & Sullivan, N=120,
Please refer to note 2 for the detailed background of the research.

Enterprises express concerns about risks associated with cloud migration, including data loss, business outages, customer attrition, application performance limitations, escalating costs, security vulnerabilities, and vendor lock-in. Among these, concerns regarding to cloud service outages or interruptions can be significantly mitigated when CSPs demonstrate strong service resilience.

It is therefore essential for enterprises to develop a comprehensive understanding of CSPs' infrastructure architectures and to establish clear criteria for assessing cloud service resilience. This knowledge empowers organizations to make informed decisions when selecting among different cloud vendors, as well as when choosing between on-premises data centers and cloud platforms.

Defining Resilience

Cloud service resilience is closely associated with service availability, reflecting the ability of systems and applications to operate continuously and deliver expected performance over time. Service availability is commonly quantified as the percentage of time a system or application remains fully operational within a given period. Uptime ratio serves as a fundamental metric for evaluating cloud resilience, offering an objective, quantifiable measure. Resilience objectives are frequently expressed as specific uptime targets.

Robust cloud resilience depends on shared responsibility between CSPs and users. CSPs achieve resilience by architecting redundant infrastructure—spanning hardware, software, networking, and operational processes—that withstands failures and enables rapid recovery from disruptions. Users, on the other hand, must design their applications with fault tolerance and efficient recovery mechanisms. CSPs support these efforts by offering tools, solutions, and best practices that empower users to strengthen the resilience of their cloud workloads.

CSPs – Cloud Service Resilience

- **Resilience of Infrastructure**
- CSPs maintain highly redundant global infrastructure—including data centers, networks, and hardware—to ensure service availability and rapid recovery from failures or disasters.

- **Resilience of Cloud Architecture**
- CSPs deploy distributed, multi-availability zone architectures with automated monitoring and failover to sustain continuous, stable operations and eliminate single points of failure.

- **Resilience of Cloud Operations**
- Through continuous health monitoring, automated management, and disaster recovery, CSPs promptly detect, address, and resolve issues, ensuring high availability and compliance with service level agreements (SLAs).

Users - Cloud Application Resilience

- **Resilience of Application**
- Users design fault-tolerant applications across multiple availability zones and Regions to eliminate single points of failure and ensure availability during cloud outages.

- **Resilience of Software**
- Users are responsible for building applications with self-healing, load balancing, and elastic scaling to enable automatic recovery and dynamic performance tuning.

- **Resilience of Application Operations**
- Users implement comprehensive monitoring, alerting, and incident response to optimize application health, swiftly detect and resolve issues, and maintain business continuity.

Cloud service resilience is fundamentally rooted in robust infrastructure. This study examines the infrastructure architecture deployed by CSPs. Achieving strong resilience in cloud services demands a systematic and highly available design, as well as rigorous deployment of the underlying infrastructure. Specifically, this entails operating multiple data centers and availability zones within each Region, enabling seamless failover between availability zones, and maintaining strict physical isolation among these components.

Cloud Infrastructure Design and Deployment are Essential to Ensuring Cloud Service Resilience

- This section explores the design and deployment of cloud infrastructure architectures by leading CSPs in China. While many providers employ similar terminologies, the precise definitions and implementations of key concepts can vary across vendors.
- We first clarify commonly used terms and highlight provider-specific interpretations, then analyze the prevailing patterns in cloud infrastructure architecture among major CSPs in China.

Main components of cloud infrastructure

In China, cloud infrastructure is characterized by several key terms, including Geography, Large Region, Region, Service Area, Availability Zone.

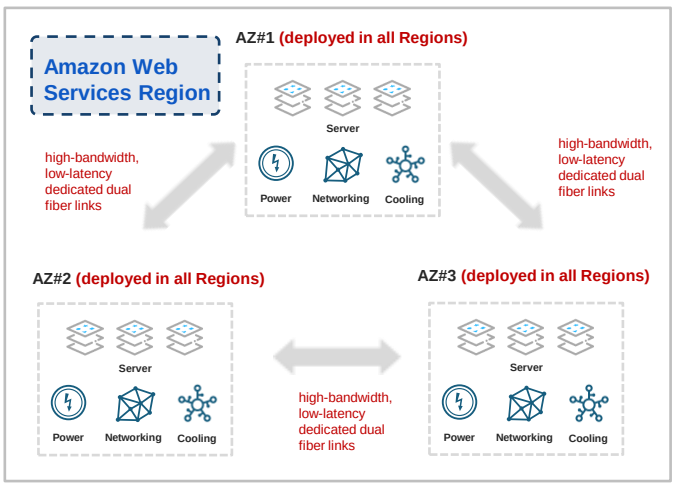
Geography: Microsoft Azure defines Geography as an area which contains one or more Regions and meets specific data residency and compliance requirements. Huawei Cloud employs a similar concept called Large Region. Amazon Web Services, Alibaba Cloud, and Tencent Cloud do not explicitly define this category.

Region: Major CSPs—including Amazon Web Services, Microsoft Azure, Alibaba Cloud, Tencent Cloud, and Huawei Cloud—broadly agree on the definition of a Region. A Region is a geographically designated area, characterized by low network latency and comprises one data center or clusters of data centers. Data centers within a Region are interconnected via redundant, ultra-low latency networks and typically share services such as elastic computing, block storage, object storage, VPC networking, elastic public IPs, and mirroring service. Regions are fully isolated from one another to maximize stability and fault tolerance. Huawei Cloud sometimes refers to Service Areas within this context.

Availability Zone (AZ): Major CSPs—including Amazon Web Services, Microsoft Azure, Alibaba Cloud, Tencent Cloud, and Huawei Cloud—adopt a consistent definition for Availability Zone. An Availability Zone consists of one physical data center or more, at distinct locations. Each AZ features independent power, cooling, networking, and security systems, and is physically separated to minimize the impact of system failures, natural disasters, and localized outages. Within each AZ, resources of computing, networking, and storage are logically segmented into multiple clusters. AZs within a same Region are interconnected via high-bandwidth, low-latency fiber-optic links to facilitate rapid communication, while AZs across different Regions remain fully isolated.

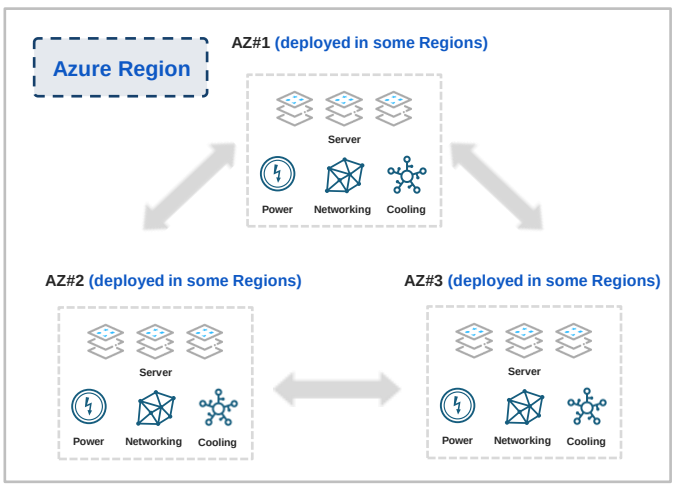
Leading CSPs in China employ varied strategies in structuring Availability Zones and data centers within each Region. Our analysis will systematically examine the cloud infrastructure architectures of Amazon Web Services, Microsoft Azure, Huawei Cloud, Alibaba Cloud, and Tencent Cloud in China, highlighting their design approaches and distinguishing features.

- **Amazon Web Services:** In mainland China, Amazon Web Services maintains an infrastructure architecture consistent with its global standards. Each of Amazon Web Services' Region is composed of at least 3 physically isolated AZs. These AZs in a same Region are interconnected by fully redundant, high-bandwidth, low-latency metropolitan fiber links.
- **Microsoft Azure:** In mainland China, Each Microsoft Azure Region comprise up to 3 AZs. Only 1 of Microsoft Azure Regions support AZs, while other Microsoft Azure Regions don't have AZs, limiting service availability and fault tolerance.
- **Alibaba Cloud:** Alibaba Cloud demonstrates notable diversity in its infrastructure deployment across mainland China. Each Region contains at least 1 AZ, with certain Regions supporting as many as 12 AZs. Approximately 43% of its Regions are provisioned with 3 or more AZs.
- **Tencent Cloud:** Tencent Cloud Regions in mainland China encompass between 1 AZ to 8 AZs. About 75% of its Regions consist 3 or more AZs.
- **Huawei Cloud:** Huawei Cloud publicly lists its Regions but does not specify the number of AZs per Region. The precise count varies and can be viewed in the Huawei Cloud console for specific resources. Each Huawei Cloud Region in mainland China includes at least 1 AZ, with each AZ corresponding to a single physical data center.



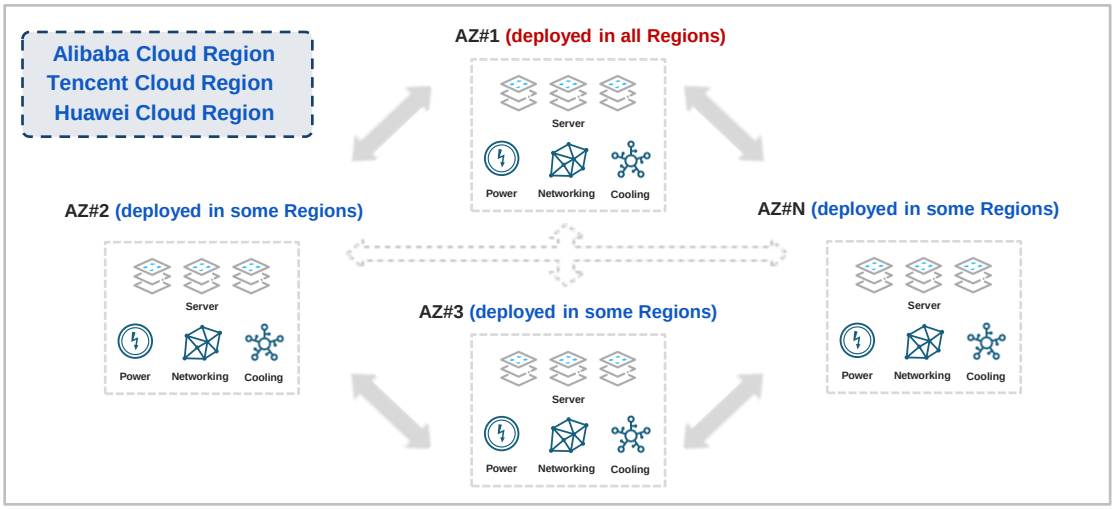
Amazon Web Services Architecture:

1. Each Amazon Web Services Region in mainland China consists of a minimum of 3, isolated, and physically separate AZs.
2. Each AZ contains one or more discrete data centers.
3. Each AZ has independent power, cooling, and physical security, enabling applications to run seamlessly across multiple AZs.
4. All AZs in an Amazon Web Services Region are interconnected with high-bandwidth, low-latency networking, over fully redundant, dedicated metro fiber links.



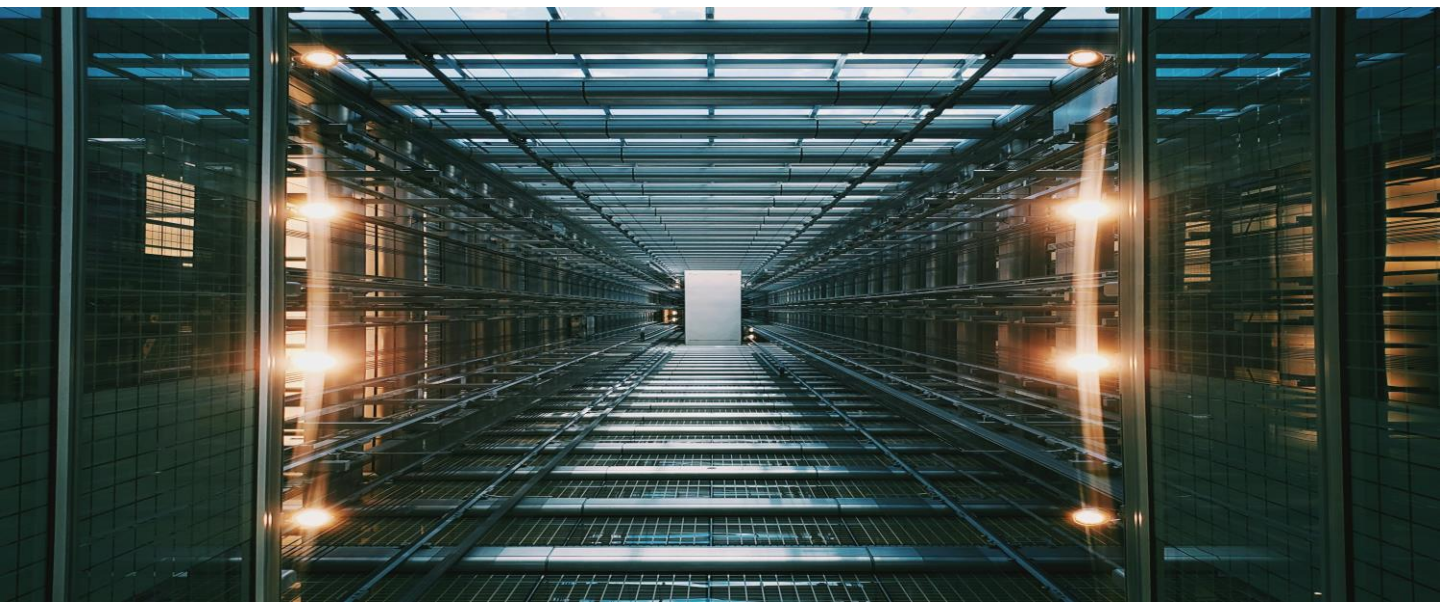
Azure Architecture:

1. Azure operates 5 Regions in China. Some of these include AZs—termed "Availability Zone Regions"—while others consist solely of data centers without AZs, referred to as "Non-Availability Zone Regions."
2. Each AZ has independent power, networking, and cooling systems.
3. Non-Availability Zone Regions rely on 1 or a few data centers to deliver basic services.



Architecture of Alibaba Cloud, Tencent Cloud and Huawei Cloud:

1. Alibaba Cloud, Tencent Cloud, and Huawei Cloud—the 3 leading domestic CSPs in China—adopt similar strategies in designing and deploying their cloud infrastructure.
2. In Regions with high data throughput and core business activities, each provider deploys 3 or more physically isolated AZs. For example, Alibaba Cloud operates more than 3 AZs in Regions in Beijing, Hangzhou, Shanghai, and Shenzhen; Tencent Cloud has over 3 AZs in Regions in Beijing, Shanghai, and Guangzhou; and Huawei Cloud maintains more than 3 AZs in Regions in Beijing, Shanghai, and Guangzhou. Each AZ consists of 1 or more data centers.
3. In Regions with lower data throughput, these CSPs generally deploy only 1 AZ or 2 AZs, applying logical isolation strategies in these cases.
4. Huawei Cloud publicly lists its Regions, while the availability zones within each Region are viewable in the Cloud Console during service configuration.



Analysis of leading CSPs' infrastructures demonstrates that a resilient cloud environment relies on thorough redundancy, resource duplication, and effective isolation.

Redundancy: The resilience of a data center cluster is determined by the number of AZs within a Region and the quality of inter-AZ communication. Each AZ must maintain resource redundancy through backup power and cooling systems, redundant power supplies for IT equipment, and multiple resource allocation pathways. Also, deploying fundamental underlying cloud services across multiple AZs and Regions can further strengthen redundancy. To achieve over 99.9% cloud service availability, it is essential to establish at least 3 AZs per Region, enabling applications to run concurrently across AZs.

Isolation: An AZ is characterized by physical separation and independence in power, cooling, networking, and other critical components. Regions that rely on a single AZ, a lone data center, or only logical isolation lack protection against failures such as power loss, cooling outages, or network disruptions. Deploying 3 or more physically isolated AZs within a Region allows for automatic incidents failover in the event of localized failures, thereby ensuring business continuity.

This analysis demonstrates that variations in cloud infrastructure design and deployment strategies fundamentally influence service resilience and availability. The ability of CSPs to effectively manage disruptions or outages depends on several critical factors.

1. **Design of Regions:** Whether regions are structured to consist Availability Zones or not.
2. **AZs and Data Centers:** The number, types, and isolation of AZs within each Region, as well as the amount of physical data centers within AZs.
3. **Infrastructure Redundancy and Interconnectivity:** The redundancy of power, cooling, and networking systems within each AZ; the availability of dedicated high-bandwidth, low-latency interconnections between AZs; and the capability to distribute user workloads across multiple physically isolated AZs.
4. **Change and Upgrade Management:** Whether system change and upgrade processes incorporate mechanisms such as phased (canary) releases and sandbox testing to maintain business continuity.

A review of major cloud outages in China underscores the necessity of robust physical isolation and multi-AZ deployment for critical services and applications. On March 29, 2023, a cooling system failure at a Tencent Cloud Guangzhou data center resulted in server crashes, disrupting core functions of WeChat and QQ—including voice calls, logins, Moments, payments, and file transfers—for approximately 12 hours. On December 18, 2022, a cooling malfunction at Alibaba Cloud's Hong Kong data center led to over 10 hours of downtime, affecting clients such as the Monetary Authority of Macau and Galaxy Macau, due to the absence of cross-AZ failover. Similarly, on June 13, 2022, a public network disruption in Huawei Cloud's Guangzhou Region caused login and trading failures for the Hithink RoyalFlush app. Collectively, these incidents demonstrate that affected cloud applications were not distributed across multiple AZs, and essential cloud service lacked sufficient physical isolation. As a result, single points of failure had widespread impacts, and existing failover mechanisms proved inadequate to maintain service continuity.

- Comparative Analysis of Infrastructure Deployment Among Leading CSPs in China

Cloud Service Provider	Minimum AZs within single Region	% of Regions with at least 3 AZs	Proportion of physically isolated AZs	Risk Level for Outages
Amazon Web Services	3	100%	100%	**Low** Each Region includes at least 3 physically isolated AZs, enabling multi-AZ application deployment, and effectively minimizes the failure impact radius.
Huawei Cloud	1	75%	92%	**Medium** Most Regions deploy 3 or more AZs but do not support application or service deployment across multiple AZs. Each AZ comprises a single data center.
Alibaba Cloud	1	42%	95%	**Medium High** Underlying cloud services lack physical isolation and are not distributed across multiple AZs, leading to extensive impact during Region-wide outages.
Tencent Cloud	1	75%	90%	**Medium High** Critical applications and underlying services apply single points deployment and face considerable challenges in achieving rapid recovery during service disruptions.
Azure	0	20%	43%	**High** Some Regions lack dedicated AZs and independent infrastructure, providing only basic cloud services.

Low risk denotes better resilience and reliability.

Low Medium Medium High High

Assessment of Cloud Service Resilience: Key Considerations

Building on initial assessments of cloud resilience based on infrastructure design and deployment, we further validate our findings through analysis of historical outages and incidents from leading CSPs in mainland China. This evaluation of performance specifically targets service continuity and failure recovery of CSPs. Our study reviews major incidents across CSPs over a 2.25-year observation period (January 1, 2023 – March 31, 2025).

Frost & Sullivan Methodology: Key Considerations in Assessing Cloud Service Resilience

Severity

Cloud outages can differ substantially in both severity and impact. CSPs generally release detailed incident reports, outlining the root cause, affected services, troubleshooting and recovery steps, final resolution, and contributing factors. Frost & Sullivan prioritized events classified as “service down,” “service failure,” or “service anomaly,” while excluding those marked as “update notifications,” “maintenance notifications,” or “adjustment notifications.”

To ensure comprehensive assessment, Frost & Sullivan evaluated the impact of each incident on both specific services and user groups, covering existing and newly provisioned instances. This approach delivers a complete overview of user experiences across all segments.

Services

When evaluating the impact of outages, our analysis focuses exclusively on incidents affecting core cloud services: computing, storage, databases, and networking. Events unrelated to these core services—such as public messaging outages, ISP blockages, or transient network fluctuations—are excluded from consideration. If a single incident impacts multiple core service categories, it is counted only once in our analysis.

Geographic Scope

CSPs specify the geographic scope of each service disruption. Our analysis is limited to incidents within Mainland China and does not include events affecting Regions as Hong Kong, Macau, or Taiwan.

Duration

CSPs document the start and restoration times of each disruption. Our methodology includes only incidents with durations exceeding 10 minutes and examines key reliability indicators—such as average and maximum outage durations, as well as annualized downtime—based on provider disclosures.

Timespan

Assessing cloud service resilience requires a long-term perspective. We analyze reliability over a period of at least 1 year to evaluate sustained performance. This study encompasses events from January 1, 2023, to March 31, 2025, capturing both long-term trends and recent developments for a comprehensive assessment.

CSPs' Disclosures

We closely scrutinize CSPs' incident disclosures, focusing on timelines, impact, root causes, remediation, resolution, reviews, and improvement strategies. Greater transparency in these disclosures reflects a provider's commitment to accountability and responsible incident management.

Reporting of Cloud Service Availability and Outage Incidents

Our analysis of cloud service outages is based on incident announcements and reports published via the official channels of each CSP. The 5 major CSPs examined—Amazon Web Services, Huawei Cloud, Alibaba Cloud, Tencent Cloud, and Microsoft Azure—all publish key details of historical outages on their official websites. However, these providers vary in the scope, transparency, and classification of their incident disclosures.To ensure consistency of our analysis, we supplement official data with additional information from publicly available online archives. The following section outlines the primary channels and key characteristics of incident disclosures among leading CSPs.

Amazon Web Services Service Health
https://health.amazonaws.cn/health/status

The dashboard presents a service list and an event log. Within the service list, users can monitor Amazon Web Services product status by date and Region. Logged-in users benefit from a personalized view that highlights events affecting their resources and flags unresolved issues. The RSS feed enables seamless integration of status data into custom IT management and analytics systems. The event log offers a continuous, detailed record of Amazon Web Services service outages over the past 12 months. Each entry includes a description of the incident, affected Regions, duration, severity, impacted services, and resolution details. For historical snapshots beyond this period, we retrieve relevant information using web archive tools.

Huawei Cloud Service Announcements
https://www.huaweicloud.com/notice.html

To identify specific service anomalies on Huawei Cloud, we consult the “Other Announcements” section of the service announcement dashboard. This section aggregates service outages, anomalies, maintenance, updates, and change notifications, requiring users to distinguish actual failures from routine events. Records are available from 2018 onward.
Each event description specifies the duration, affected Regions, and impacted services, but does not provide information on incident resolution.

Alibaba Cloud health dashboard
https://status.aliyun.com/#/historyEvent

The Alibaba Cloud health dashboard delivers real-time service status and a historical event log, both accessible via RSS. Each event entry specifies the start and resolution times, affected Regions, and impacted services, but does not provide details on resolution steps. The dashboard archives events for the past 12 months; incidents prior to this period are tracked using web archives.

Tencent Cloud health dashboard
https://status.tencentcloud.com

Tencent Cloud's health dashboard also offers RSS access for real-time status updates. The historical events panel documents incidents over the last 12 months, detailing affected services, Regions, and duration. Notably, Tencent Cloud provides in-depth analysis of incident progression and resolution. For earlier snapshots, relevant information is retrieved through web archive tools.

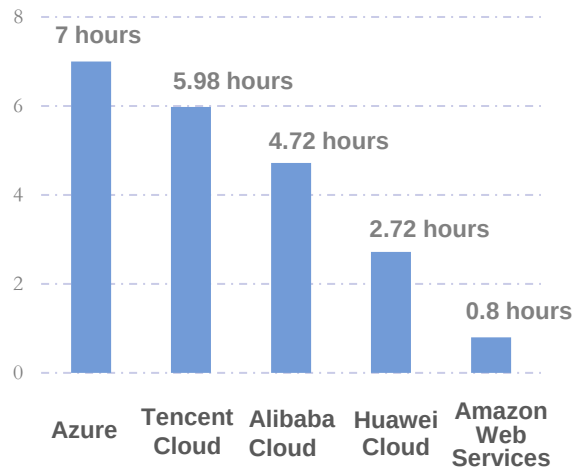
Microsoft Azure Status History
https://azure.status.microsoft/zh-cn/status/history/

Microsoft Azure maintains a comprehensive, global record of cloud service anomalies in its status history. Since November 20, 2019, Azure has publicly documented detailed incident timelines, resolutions, and post-incident reviews (PIRs), retaining each PIR for 5 years. For every event, Azure discloses the duration, affected services and Regions, root cause analysis, improvement strategies, and user guidance, demonstrating a high level of transparency and completeness.

Service Resilience among Cloud Service Providers

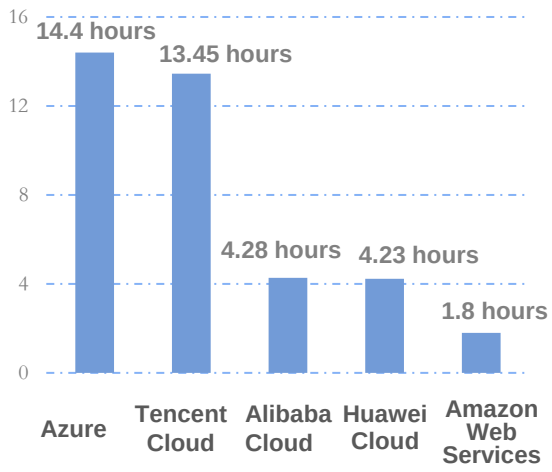
- An analysis of cloud services availability from the 5 leading CSPs during the selected period reveals that Amazon Web Services demonstrated superior resilience in China compared to Huawei Cloud, Alibaba Cloud, Tencent Cloud, and Microsoft Azure. Amazon Web Services experienced shorter anomaly and outage durations, resulting in higher overall service availability. The following diagrams and tables provide a detailed breakdown of these findings.
  - **Analysis Timeframe:** January 1, 2023 - March 31, 2025.
  - **Regions:** Incident impacted any Region falling within mainland China
  - **Incidents included:** Outage or disruption to core services

Average Duration of Cloud Service Incidents on an Annual Basis[3],
hour



3. Calculation: Aggregate duration of incidents ÷ the number of years

Average Duration of Cloud Service Incidents on a Region-wide Basis[4],
hour



4. Calculation: (incident #1 * number of Regions affected +……+ incident #N * number of Regions affected) ÷ total number of Regions

To initiate our analysis, we compared the average annual duration of cloud service anomalies or outages across the 5 CSPs. Huawei Cloud, Alibaba Cloud, Tencent Cloud, and Microsoft Azure each recorded average annual incident durations exceeding 2 hours. Amazon Web Services was the only provider with average annual anomaly or outage durations below 1 hour, achieving 99.9909% availability[5]—significantly surpassing its SLA commitment for mainland China. Among the other providers, Huawei Cloud led with 99.9689% availability, meeting its SLA commitment in China.

5. Availability is calculated as: (the number of hours the core cloud services remain operational ÷ total amount of hours) × 100%.

Given that a greater number of Regions increases the likelihood of cloud service anomalies or outages, we also analyzed the average incident duration per Region for each CSP during the study period. Taking a comparative view of the "annual average" and "Region-wide average" incident durations among CSPs, we observe that Amazon Web Services consistently maintained a Region-wide average of under 2 hours. Alibaba Cloud also exhibited strong performance in terms of Region-wide average. In contrast, both Tencent Cloud and Microsoft Azure showed significantly weaker Region-wide averages—with incidents duration twice longer than their respective annual averages. Similarly, Huawei Cloud also performed less favorably in the Region-wide average metric.

Cloud Service Provider	Frequenc: # of incidents	Aggregate duration of incidents	Average duration	Longest duration for a single incident	Availability (Uptime) Percentage
Amazon Web Services	1	1.8h	1.8h	1.8h	99.9909%
Huawei Cloud	4	6.12h 3.4X more than Amazon Web Services	1.53h	4h	99.9689%
Alibaba Cloud	5	10.62h 5.9X more than Amazon Web Services	2.12h	4.7h	99.9461%
Tencent Cloud	2	11.45h 6.3X more than Amazon Web Services	5.73h	10h	99.9419%
Azure	2	15.75h 8.7X more than Amazon Web Services	7.88h	13.5h	99.9201%

The table above provides a detailed comparison of cloud service resilience among leading CSPs in China during the study timeframe. Amazon Web Services demonstrated exceptional resilience, recording only a single service anomaly and achieving 99.9909% availability. Compared to Microsoft Azure—the other major non-domestic provider—Amazon Web Services has made a stronger commitment to service stability in China, with infrastructure deployment that offers greater resilience than that of domestic providers.

Huawei Cloud ranked second in resilience after Amazon Web Services. Although Huawei Cloud experienced a higher frequency of incidents, each event was shorter in duration than those of Alibaba Cloud, Tencent Cloud, and Azure, indicating robust monitoring and rapid response.

Alibaba Cloud's cumulative incident duration exceeded that of Huawei Cloud by 4.5 hours. In 2023, two major global outages impacted Alibaba Cloud's services in China, revealing vulnerabilities in its underlying service continuity.

Among domestic providers, Tencent Cloud reported the longest incident duration—1.8 times that of Huawei Cloud. Despite having the same number of Regions (8) and multi-availability zone Regions (6) as Huawei Cloud, Tencent Cloud's lower availability likely results from insufficient physical isolation and less robust infrastructure deployment.

Microsoft Azure operates the fewest multi-AZ Regions in China (just 1), with most deployments functioning as logical data centers. One of its two major disruptions affected all Azure Regions in China, highlighting infrastructure limitations. However, Azure excels in incident reporting transparency, consistently providing detailed root cause analysis, improvement plans, and user guidance.

In summary, the service resilience of leading CSPs in China closely aligns with their infrastructure deployment strategies, as detailed in earlier analyses. Both assessments rank the providers identically from highest to lowest resilience: Amazon Web Services, Huawei Cloud, Alibaba Cloud, Tencent Cloud, and Microsoft Azure. A 2024 Frost & Sullivan survey found that each minute of digital infrastructure downtime can cost enterprises anywhere from thousands to tens of thousands of RMB, depending on industry and company size.

These findings underscore the importance of robust infrastructure design and deployment, including comprehensive redundancy and isolation measures. Such strategies are essential for maximizing availability, minimizing outages, and reducing business disruption.

What it means: Amazon Web Services demonstrated exceptional resilience in its cloud services across mainland China, distinguishing itself as the only provider to consistently exceed 99.99% availability during the study period. Furthermore, Amazon Web Services's total downtime was less than 1/5 of the average downtime recorded by other providers.

Amazon Web Services's exceptional performance stems from its globally consistent infrastructure design and deployment approach. This encompasses multiple strategic layers—including availability and redundancy engineering, robust network connectivity, multi-AZ operations, and synchronous data replication—all seamlessly integrated into both the architecture and user workflows to ensure resilience.

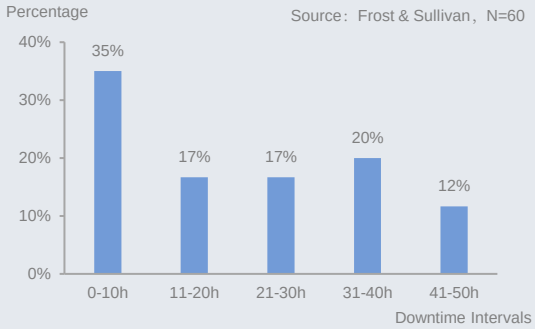
1. Amazon Web Services leads the industry by prioritizing cloud infrastructure availability. It is the only provider in the China mainland to deploy at least 3 physically isolated AZs within each Region. Each AZ features fully independent redundancies for power, cooling, and network systems, ensuring complete isolation against infrastructure and network failures. This design fully realizes the benefits of a multi-AZ architecture.
2. Amazon Web Services transparently discloses the physical distances between its AZs, maintaining separations of up to 100 kilometers within a Region. This approach prevents fault propagation between AZs while enabling efficient inter-AZ communication to support real-time data synchronization. Deploying applications across multiple AZs thus provides users with robust continuity assurances.
3. All AZs within an Amazon Web Services Region are interconnected via high-bandwidth, low-latency networks over fully redundant, dedicated metro fiber. This infrastructure ensures high throughput and minimal latency, facilitating efficient synchronous replication between AZs. Each AZ is also connected to Tier 1 internet service providers through 2 independent transit centers, ensuring resilient and redundant public network access.
4. Amazon Web Services provides users with clear visibility into Regions, resources within AZs, fault domains, and AZ architecture, empowering them to leverage multiple Amazon Web Services services to design tailored resilience strategies. Compared to domestic CSPs, Amazon Web Services offers greater flexibility for multi-AZ application deployments and better control over AZ selection.

How does the resilience of on-premises data centers measure up against that of cloud?

China's intelligent industry is expanding rapidly, driving strong demand for computing resources. As a result, more enterprises face the choice of building on-premises data centers or migrating workloads to the cloud. According to Frost & Sullivan, around 75% of Chinese companies are willing to store non-core and auxiliary business data on the cloud, while only about 10% are comfortable storing core business data on the cloud. Although China's cloud adoption over the past 3 years has kept pace with global trends, many enterprises remain cautious about cloud service resilience, often viewing on-premises data centers as more reliable for ensuring business continuity.

In the first quarter of 2025, Frost & Sullivan surveyed 120 Chinese enterprises to assess the reliability of local data centers versus cloud services. Respondents were grouped based on their use of either local data centers or cloud platforms. The survey revealed that on-premises data centers experienced an average annual downtime of 22 hours, corresponding to an average annual availability of approximately 99.7%. In contrast, leading CSPs reported an average annual downtime of only 4.2 hours, achieving over 99.92% availability. These results clearly demonstrate that cloud offer significantly higher availability and resilience compared to on-premises data centers.

Average annual service downtime at on-premises data centers



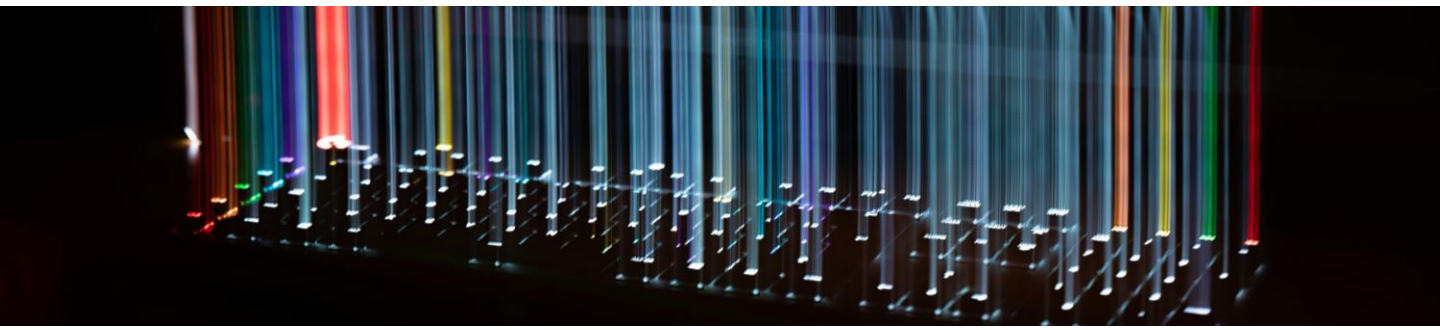
On-premises data centers' average annual service downtime
22 hours

VS.

CSPs' average annual service downtime
4.2 hours

Beyond the core questions, our survey examined enterprise perspectives on future investments in on-premises data centers. Findings revealed that over 90% of companies operating on-premises data centers are uncertain whether increased investment would improve service availability or continuity. Among these respondents, only about 15% plan to expand their on-premises data center investments over the next 3 years, while roughly 50% are considering migrating workloads to the cloud within the same timeframe. The remainder intend to maintain their current infrastructure.

Among enterprises already using cloud services, approximately 70% reported plans to migrate additional workloads to the cloud in the coming years.



Steps Enterprises Can Take to Improve Resilience of Cloud Workloads

While the resilience of cloud services ultimately hinges on how CSPs design and deploy their infrastructure, enterprises can proactively adopt strategies to strengthen the continuity of their cloud operations.

- Enterprises should assess the risk tolerance of different type of workloads and select resilience strategies accordingly—whether at the AZ or Regionl level. For non-critical workloads, deployment within a single AZ may be cost-effective. Production workloads should be distributed across multiple AZs, and mission-critical workloads should adopt architectures spanning multiple AZs and Regions.
- Embracing cloud-native architectures and tools significantly enhances recovery efficiency. Simply migrating applications without leveraging cloud-native automation and processes limits the cloud's benefits. Even multi-Region disaster recovery setups may struggle with rapid recovery if not fully cloud-native. Therefore, we recommend optimizing architectures using cloud-native tools such as Infrastructure as Code (IaC), which increases agility and enables swift failover to backup Regions during outages.
- Enterprises should verify whether their workloads support automatic failover. When applications are deployed within a single AZ due to latency sensitivity, responsibility for failover and offsite data backups typically falls on the user. CSPs generally enable automatic failover only in multi-AZ deployments.
- When selecting cloud service locations, enterprises often prioritize customer experience by deploying applications close to users. However, to ensure business continuity and system resilience, it is critical to balance customer experience with robust deployment strategies. Concentrating all cloud workloads in a single Region or AZ can undermine recovery capabilities during unexpected failures.
- For workloads already running on the cloud, enterprises should regularly reassess resilience requirements to determine if workload criticality or risk tolerance has changed, and proactively collaborate with CSPs for necessary calibration.

The Last Word

When choosing a cloud provider, enterprises should carefully evaluate factors such as availability, cost efficiency, agility, operational complexity, and—most importantly—resilience, which is fundamental to service reliability. Organizations require uninterrupted access to business data and workloads at all times. While some may attempt to ensure this by building on-premises data centers, comparisons over a long period indicate that cloud infrastructures generally offer greater robustness and fault tolerance.

We recommend that enterprises thoroughly assess CSPs' infrastructure—including the number and design of Availability Zones and Regions, redundancy configurations, and historical resilience performances such as outage records. This comprehensive evaluation enables organizations to select cloud solutions that best align with specific requirements.

ABOUT US

With a team of growth coaches based in 45 global offices, we have mastered the art of identifying Growth Opportunities in hundreds of sectors using a powerful understanding of how value chains operate on a global level. Frost & Sullivan's innovative go-to-market strategies and proven implementation Best Practices have helped transform the business models of some of the world's leading companies through our Growth Pipeline as a Service, which makes it easy for their customers to create and implement a continuous flow of Growth Opportunities.

Frost & Sullivan is at the center of an ecosystem of best practice coaching, executive peer support communities, and growth-oriented content that is singularly focused on reshaping the world through managed growth.