ZIdentity Resilience Guide

Overview of the ZIdentity High Availability Architecture

AUTHOR

HARNEET SINGH

WHITEPAPER

Overview

Trust is paramount when adopting cloud services to support mission-critical systems and capabilities. It's essential to have confidence that your cloud provider is not only available and scalable but also resilient, reliable, and as dedicated to security and compliance as your organization. This becomes even more crucial in the context of Identity and Access Management (IAM). ZIdentity is the centralized identity platform for securing access to all your applications.

Zscaler ZIdentity is designed with trust at its core. We have carefully crafted the technology, streamlined processes, and assembled the expertise needed to offer our customers a suite of best-in-class solutions that adhere to the highest standards. ZIdentity is:

- **Built for Web Scale:** Seamlessly scaling up or down to meet your needs.
- **Availability:** Architected for negligible downtime, with no maintenance windows required.
- **Secure:** Rigorously audited with a best-in-class security team.
- **Constantly Evolving:** Enabling new capabilities and rapid innovation, while insulating your organization from the ever-changing IT landscape.

Resilient architecture

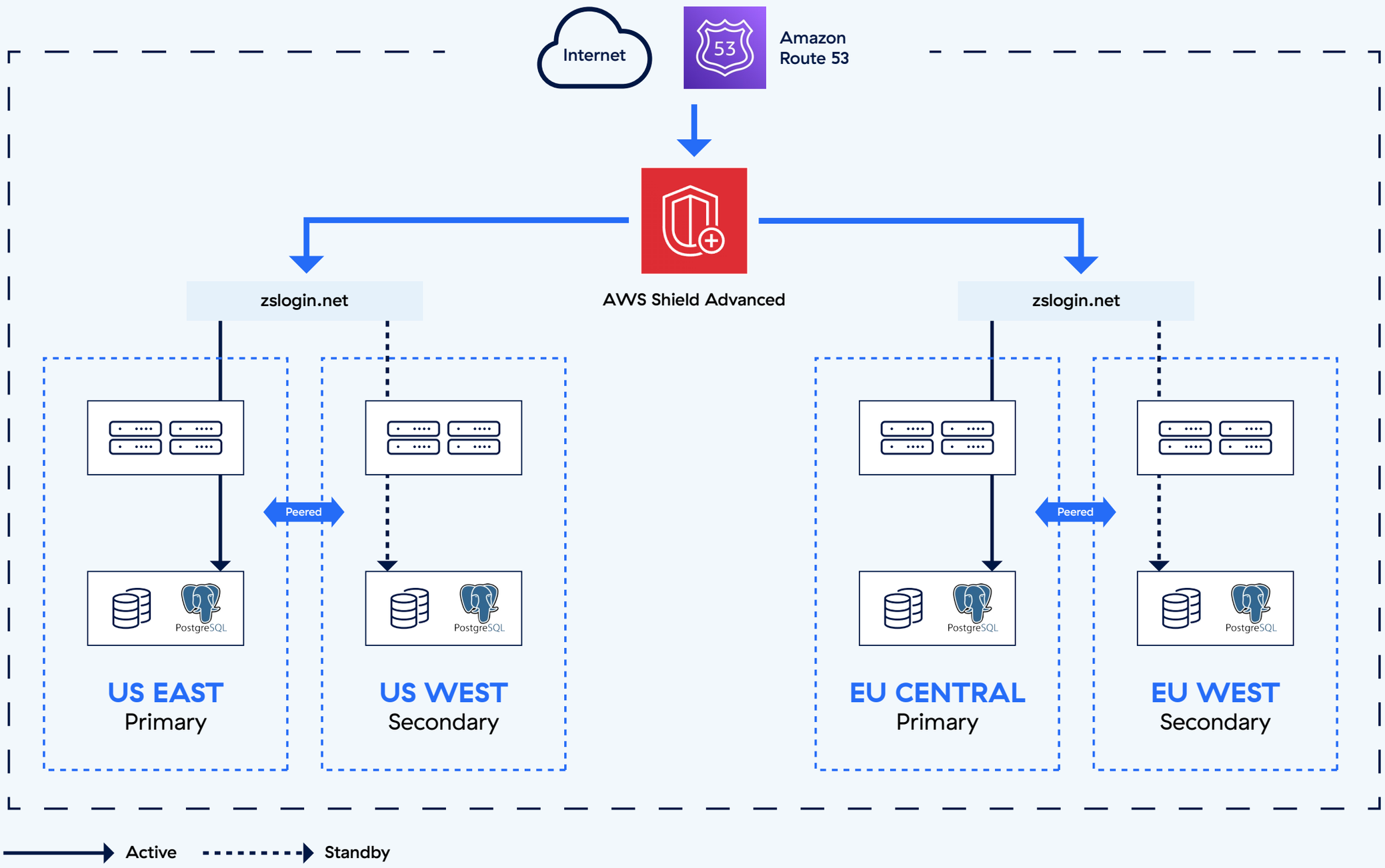
The key feature of the ZIdentity architecture is its fully multitenant design, where customers share a common underlying environment. This approach enables significant economies of scale, allowing Zscaler to build a highly robust infrastructure with advanced redundancy, comprehensive monitoring, and streamlined processes.

Each ZIdentity environment is a self-contained unit, which encompasses all the components needed to operate a ZIdentity deployment. These environments are capable of functioning independently, forming a critical part of our availability strategy by minimizing the impact of potential outages to customers.

Within an instance, the architecture is organized into three primary tiers:

- **Front-End Tier:** Includes proxy load balancers, ingress and firewall services to manage traffic efficiently.
- **Application Tier:** Houses the core software that powers ZIdentity.
- **Database Tier:** Features optimized database services tailored for high performance.

The entire services is hosted on Amazon Web Services (AWS), utilizing multiple availability zones and geographically distributed regions. This design ensures exceptional scalability, throughput, and reliability — meeting the demands of high-performance operations while addressing key GDPR and data sovereignty requirements.



Stateless Services

All components, with the exception of databases, are functionally stateless. Consequently, any server above the database layer can process any request. This means that all system components can be scaled as needed by launching additional services in AWS. This flexibility enables ZIdentity to dynamically respond to fluctuations in demand and seamlessly reroute traffic from failed services to other active services. Requests are queued and handled by available resources, ensuring that failures are transparent to users. If an application server fails, authentication requests are automatically forwarded to a working server for retry.

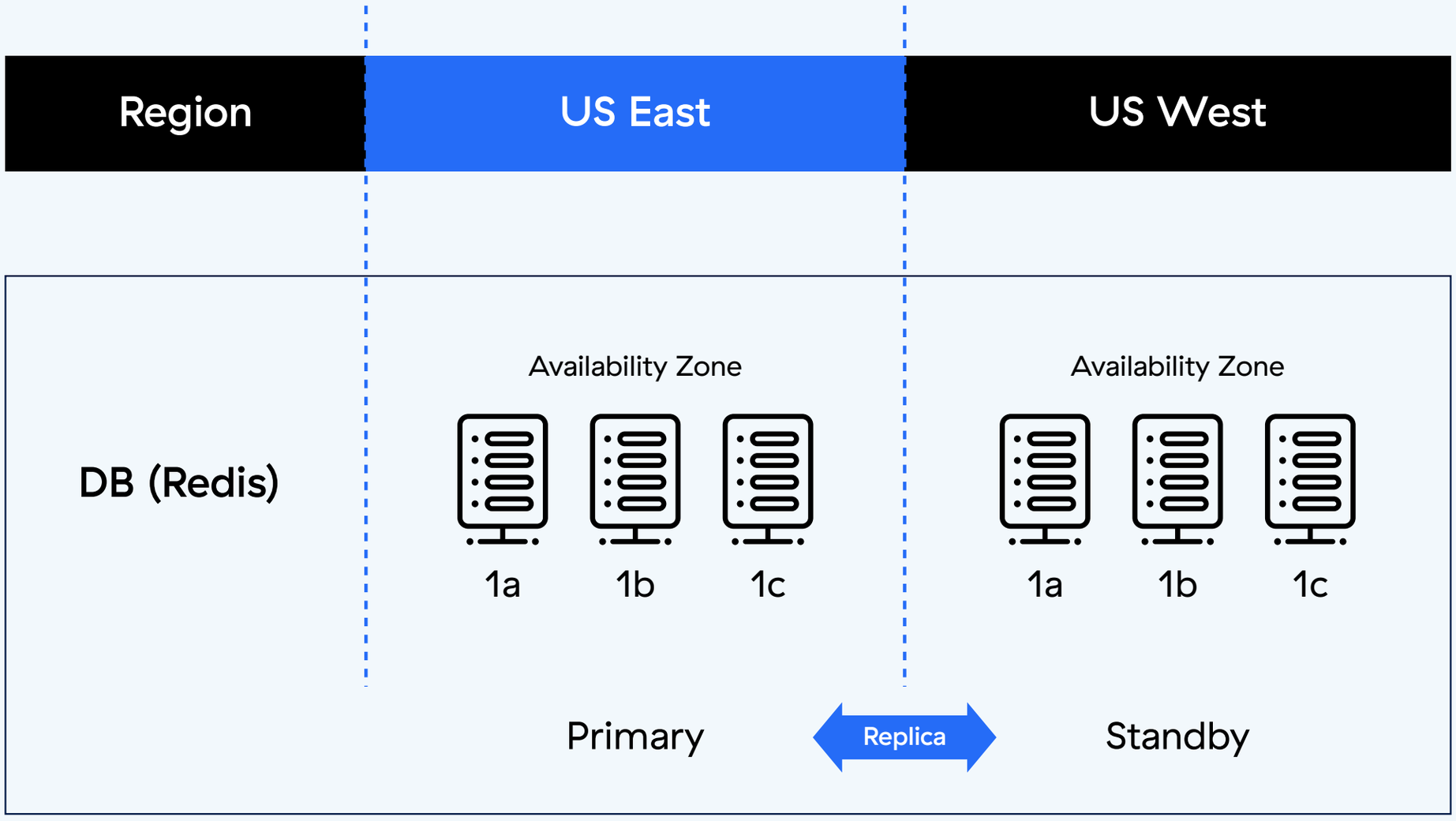
Stateful Services

ZIdentity services also host stateful services designed to support asynchronous message processing. These stateful services are deployed and configured in a manner that allows them to dynamically scale in terms of compute, storage, and network bandwidth, ensuring optimal performance even during high-demand scenarios.

This scalability ensures that services can handle varying workloads seamlessly, maintaining availability and operational continuity. By efficiently managing resource allocation, these stateful services enable the system to function without interruptions, even in complex, dynamic environments, thus enhancing reliability and resilience.

Databases optimized for functionality

Although stateless components above the data tier can be added or removed at any time, this does not apply to the database layer. Database services must maintain state and manage transactions to ensure customer data remains intact and is not corrupted during outages. We use a highly available configuration eliminating any single point of failure. The instances are spread out across the availability zone where if instances go down, the other instances would take over and new instances would be spun up. In case the primary region goes down, a secondary region is promoted to act as Primary.



No single points of failure

SERVICE DISRUPTIONS

Our platform is built with a strong commitment to high availability, and while occasional challenges may arise, we see each as an opportunity to learn, adapt, and improve. We proactively address potential issues and implement robust measures to enhance reliability. By continuously refining our infrastructure and processes, we aim to deliver an exceptional experience for our users, embracing the ongoing journey toward perfection.

Team Readiness and Support

Recovering the service during an incident also depends on the efforts of Zscaler team members who are prepared to provide support. To address this, Zscaler takes proactive measures to ensure that critical teams are geographically distributed, reducing the impact of regional disasters. Each team is equipped with diverse communication channels, enabling effective coordination even in scenarios involving the loss of production and corporate services. This ensures that engineering teams remain focused on restoring the service promptly and efficiently.

Data backup

Replication across multiple availability zones and regions ensures that ZIdentity is able to function as a highly available service, but it does not provide adequate protection in the event of a catastrophic failure of AWS infrastructure or data corruption. Therefore, Zscaler also takes frequent database snapshots and stores them in a read-only, protected account. These encrypted, protected backups serve as a failsafe mechanism that enables Zscaler to restore the service.

Platform with Continuous Availability

UNINTERRUPTED AVAILABILITY WITHOUT SCHEDULED DOWNTIME

At Zscaler, we believe consistent availability should never compromise the ability to deliver continuous innovation. Historically, balancing high availability with ongoing service enhancements has been a challenge. However, we've overcome this by implementing a canary release architecture and a fully automated testing and deployment process. This methodology allows us to deliver continuous innovation without planned downtime or maintenance interruptions.

Every code change committed undergoes a comprehensive series of functional and security tests. Once the code successfully clears this rigorous automated process, it is deemed ready for deployment. The operations team facilitates a smooth rollout of the fully tested code to developer environments, where it undergoes further testing. Updates are then deployed to a beta environment, allowing customers to conduct their own acceptance testing. Following a designated preview period, the updates are moved to production.

This thorough and carefully orchestrated approach enables Zscaler to maintain a highly reliable and innovative platform. By eliminating planned downtime, we ensure our customers can consistently rely on our service for their critical needs.

Real time monitoring

Disasters are rarely the result of a single event; instead, they typically stem from a series of interrelated events that must occur at the same time. Early identification and intervention in this chain of events is essential for a robust high-availability strategy, which is why Zscaler emphasizes proactive service monitoring.

The primary defense against unforeseen issues is comprehensive instrumentation and monitoring across all system components, categorized into external monitoring & alerts and internal monitoring & alerts.

For external monitoring, Zscaler utilizes several services with globally distributed test agents to continuously monitor application performance and provide real-time alerts. This offers a constant stream of reliable data on the service's operation. Additionally, we deploy internal monitors to detect system issues, but their main advantage is that they help us pinpoint which specific services are encountering problems.

These internal monitors are more sensitive than external ones and often detect issues before they impact site performance or availability. Zscaler integrates internal monitoring across all subsystems and software components to ensure full visibility.

Rapid disaster recovery

Zscaler ensures secure access to critical customer data, which must remain available at all times. To meet this need, Zscaler's architecture is designed for high resilience and continuous availability.

This approach ensures that even during incidents, your operations remain minimally affected. While data backups are vital, a disaster is not the time to test your recovery capabilities. That's why Zscaler performs disaster recovery testing on a regular basis. These exercises ensure the team is fully prepared to handle emergencies effectively, and all data can be restored without issue.

By integrating these practices, Zscaler maintains the integrity and availability of critical customer data, ensuring continuous and secure access even in the face of unexpected disruptions.

Final Thoughts

The Zscaler ZIdentity high availability architecture exemplifies a robust, scalable, and secure platform designed to meet the demanding requirements of modern Identity and Access Management. By leveraging a multi-tenant architecture, stateless services, optimized databases, and proactive disaster recovery measures, Zscaler ensures unparalleled reliability and minimal disruption. Continuous availability is supported by innovative deployment strategies, comprehensive monitoring, and geographically distributed team readiness, providing organizations with a seamless and dependable IAM solution.

ZIdentity's commitment to eliminating single points of failure, maintaining data integrity, and delivering real-time responsiveness reaffirms Zscaler's dedication to securing mission-critical operations while adapting to evolving business needs. This architecture underscores Zscaler's pledge to empower organizations with trust, innovation, and confidence in their cloud-based identity services.

+1 408.533.0288 **Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134** **zscaler.com**

Zero Trust Everywhere