How Duo Security Delivers
Service Reliability

Published January 29, 2021

Building a Reliability Culture

The Duo software as a security (SaaS) platform protects our customers' users, devices, applications and data. Recognizing the critical nature of these capabilities, we invest heavily in architecting and operating an ultra scalable and highly reliable service.

Service reliability is of utmost importance across all levels of the Duo organization and we've designed systems, processes and team structures to reflect this. Multiple Engineering teams focus exclusively on software quality and service reliability. We build frameworks, tools and automation that ensure best practices are consistently applied across all aspects of the Duo service.

We measure everything possible, collecting operational metrics surrounding every interaction with the Duo platform. This enables our data-driven approach to managing and maintaining the Duo service, feeding systems that dynamically scale the platform and proactively call the team to action when necessary. Additionally, we aggregate this information and use it to establish key performance indicators that track overall platform health.

We are constantly improving our systems and processes, gradually and consistently implementing changes that improve reliability, scalability and security of the Duo platform. We architect our systems to gracefully handle infrastructure or external service failures with no impact to Duo customers and to seamlessly scale to meet customer needs.

Our overarching goal is to provide a service that is secure, always available and consistently performs well. We take this responsibility very seriously and are proud of our multi-year track record of excellent service reliability.

Providing Highly Available Services

The Duo platform leverages multiple vendors and service providers, including Amazon AWS which provides many of the underlying infrastructure components. These infrastructure components are backed by real-time replication between the multiple Availability Zones (physical data centers) that make up an AWS Region (grouping of multiple data centers in a geographic region).

This means that a failure affecting any single data center will not affect the availability of Duo services. Additionally, Duo also replicates customer data in real time to at least one additional AWS Region to be used in the event of an AWS service outage affecting an entire region. A high-level overview of this technology can be found below:
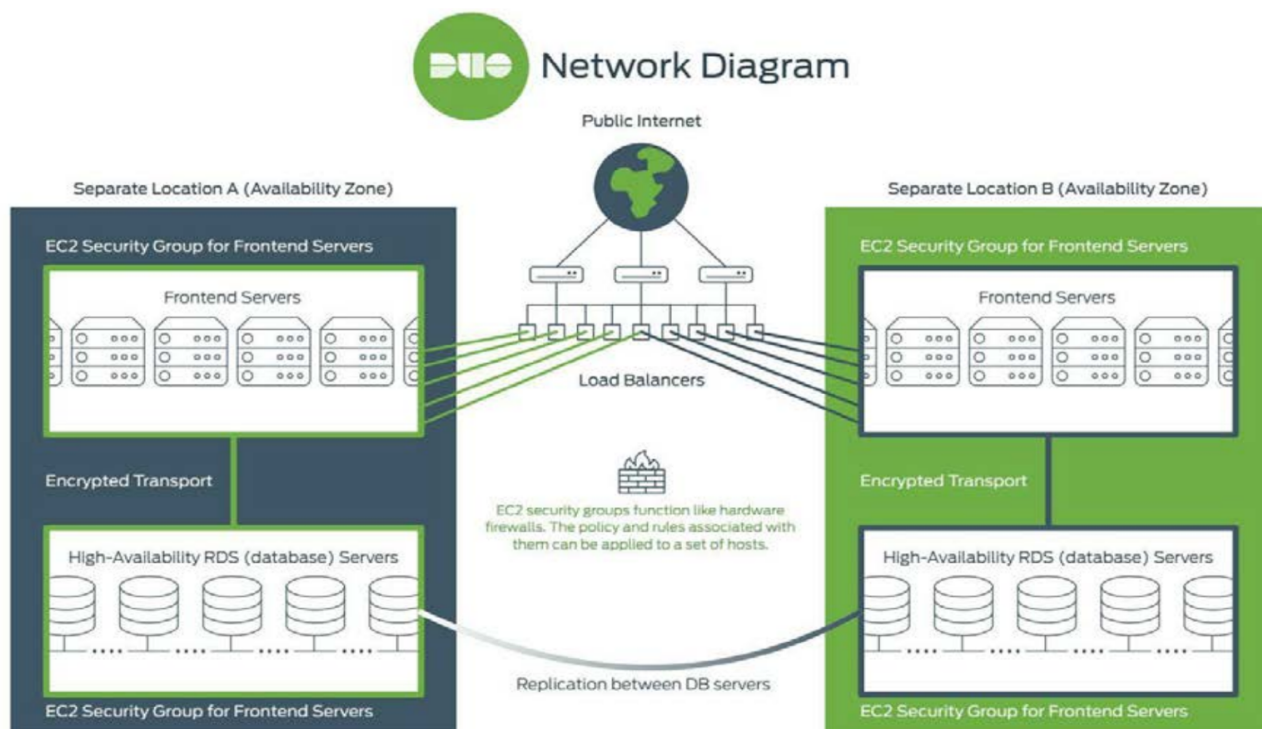
This architecture has been designed with redundancy at each tier, and is configured to automatically recover in the event of a failure, without requiring manual intervention. This automatic recovery takes advantage of functionality provided by AWS, as well as fault tolerance capabilities we've built into our platform.

In the event of an incident not automatically recovered from, Duo's Production Engineering team is engaged to assist in recovery efforts. Our Production Engineering team is responsible for the availability of this infrastructure and Duo services through monitoring and alerting, as well as 24/7 on-call support.

Additionally, Duo has empowered customers to define their own criteria as to how they'd like the service and associated integrations to respond in the event of a service-impacting incident. See our Guide to Business Continuity Preparedness for more information.

Also visit the Duo Status Page for a real-time and historical view on Duo service status and to sign up for incident alerts.

Scaling for Growth

Duo utilizes numerous independent clusters or "deployments" of our technology to both scale to support customer growth, and to ensure that the impact of a failure in any one deployment is minimized. Each of these deployments is built from the same template and managed using our automation technology to ensure they continue to share the same high availability capabilities. Because these deployments are built using automation on top of cloud technologies, the underlying infrastructure behind each deployment is refreshed on a continuous basis.

As part of scaling for customer growth, we're able to add additional capacity at each tier (application, database, etc.) of each deployment in a highly-tested and repeatable fashion, using this same automation technology. Identifying the need for this sort of scaling well in advance of warning thresholds is the responsibility of our SRE (Site Reliability Engineering) teams, with data being constantly evaluated and incorporated into capacity management procedures.

As Duo does not maintain any sort of standing maintenance window that allows for customer downtime, we've designed our architecture in such a way that maintenance can be performed against any portion of the infrastructure without causing customer impact. These sort of infrastructure management activities are planned and executed across all Duo deployments on a consistent basis, transparently to our customers.

Delivering Value to Our Customers

Duo releases software every two weeks to give customers access to new features and service improvements.

We've implemented a rigorous testing program that seeks to ensure these software releases are free of regressions through the use of automated unit and integration tests, as well as a full suite of acceptance tests conducted by our Engineering teams as part of every release.

In addition to our testing processes, we practice a rolling release process to slowly release these changes to our fleet of deployments. Our Engineering team has built processes allowing these types of changes to be made in a gradual and automated fashion, and these processes are exercised regularly as Duo releases code. Duo's own deployment is the first to receive updates, providing additional real-world validation of each release before it reaches our customers. From there, we begin rolling out these updates to customer deployments, and continue to do so during a two-week cycle.

The following table shows a high level overview of this process:

Week 1	Week 2	Week 3	Week 4
Internal Development + Automated Testing	Internal Development + Automated Testing	Internal Validation + Production Releases + Ongoing Monitoring	Production Releases + Ongoing Monitoring

This rolling release process is designed to minimize the impact of unexpected software regressions or other related issues, and allow us to resolve these issues before broader release to our customers. In the event that a software regression or other issue is discovered, further releases are halted until a fix can be developed and integrated into that release. The ordering of deployments as part of this rolling release process is frequently evaluated and adjusted to ensure no single deployment is disproportionately affected by software regressions, as infrequent as they may be.

In short, we're committed to providing a service that our customers and their end users can always depend upon. We honor this commitment by investing in our people, processes and technology to ensure that the Duo service remains available and secure at all times.

Additional Resources

You can find historical release notes here.

You can subscribe to future release note updates following these instructions.