Advanced High Availability Architecture
Delivering performance, scalability, and availability with the Now Platform
Introduction

Organizations rely on access to IT and business data and services for their continued operation and success.

This document provides an overview of the ServiceNow® Advanced High Availability (AHA) architecture—a key element in delivering an enterprise-grade cloud service. The unique, multi-instance architecture not only meets but exceeds stringent requirements surrounding data sovereignty, availability, and performance.
## Table of Contents

Advanced High Availability ......................................................... 4
Multi-instance architecture ........................................................ 4
Global data center pairs............................................................. 5
Real availability dashboard......................................................... 5
Performance and scalability....................................................... 6
Critical resources........................................................................ 6
Backup and recovery................................................................. 6
Transfer and failover................................................................. 6
Conclusion ................................................................................. 7
Advanced High Availability

ServiceNow’s data centers and cloud-based infrastructure are designed to be highly available with redundant components and multiple network paths to avoid single points of failure.

At the heart of this architecture, each customer application instance is supported by a multi-homed network configuration with multiple connections to the internet from different providers and with redundant power sources. Production application servers are load balanced within each data center and are replicated in near real-time with the paired data center within the same geographic region.

We leverage AHA for customer production instances for the following purposes:

• Prior to executing maintenance, ServiceNow can proactively transfer operation of a customer instance from one data center to the other. The maintenance can then proceed without impacting service availability.
• In the event of the failure of one or more infrastructure components, service is restored by transferring the operation of the affected instance to the other data center.

With this approach, the transfer between active and standby data centers is regularly executed as part of our standard operating procedures. This ensures that when it is needed to address a failure, the transfer will be successful and service disruption is minimized.

Multi-instance architecture

Instances of the Now Platform® are deployed on an advanced, multi-instance architecture that provides separate application nodes and database processes for each customer. This ensures that there is no possibility of co-mingling of customer data, even between instances assigned to the same customer, unlike multi-tenant architectures where a shared database is used.

Each instance runs its own application logic and database processes, meaning that an instance does not have to be on the same version or upgraded at the same time as other customers instances. Customers can choose to upgrade their instances on a schedule that best meets their needs and compliance requirements. No downtime is necessary for upgrades.
Global data center pairs

ServiceNow’s data centers are arranged in pairs. There are ten support centers and eleven data center pairs, spanning five continents: Asia, Australia, Europe, North America, and South America.

All customer production data is stored in both data centers and kept in sync using real-time database replication. Both data centers are active at all times, each with the ability to support the combined production load of the pair. ServiceNow maintains continuous, asynchronous replication from the database in the current primary data center (read-write) to the secondary data center (read-only).

Real availability dashboard

We provide a real availability dashboard that displays availability information for all of a customer’s instances, providing a true measure of customer availability.

For further details, see the Availability section in Securing the Now Platform.
Performance and scalability

Our cloud scales to meet the needs of the largest Global 2000 enterprises, with tens of thousands of customer instances operating in our globally distributed data centers. All instances are deployed on a per-customer basis, allowing the multi-instance cloud to scale horizontally to meet each customer's performance needs.

Customer instances perform an aggregate of tens of billions of full-page transactions every month, and customers using the ServiceNow configuration management database (CMDB) as their single system of record may manage tens of millions of configuration items (CIs).

Critical resources

ServiceNow is responsible for managing its environment, the supporting infrastructure, and vendor relationships. As part of these responsibilities, ServiceNow's site reliability engineering (SRE) center employs a follow-the-sun model that provides continual security, operational monitoring, and support of the ServiceNow environment and infrastructure. ServiceNow rotates operations and technical support daily in North America, the UK, India, Australia, Netherlands, Japan, and Ireland.

Critical system resources, including DNS, email, ServiceNow's cloud operations systems, and customer service system are operated in high availability configurations in a minimum of two data centers. None of these resources rely upon ServiceNow's internal corporate IT infrastructure.

ServiceNow uses AHA for its own development systems, including managing source code control and the software build process that are also hosted at the production data centers. This ensures the highest continuity for our developers, enabling them to continue to develop and support the application without requiring physical access to ServiceNow offices.

The AHA architecture uses the same transfer process for preventive maintenance and recovery from natural disasters. This approach eliminates the need for a yearly disaster recovery test and creates a practiced transfer event during the performance of normal maintenance.

Backup and recovery

Backups of customer instances are also taken primarily for logical restore purposes, such as the case of a critical customer configuration error or accidental data deletion.

Full backups are performed every seven days direct to disk, and are retained for 28 days, with differential backups taken every 24 hours. Backups are stored in the same data centers where the data resides, with production instances backed up in both data centers in the pair. Sub-production instances (commonly used for testing and development purposes) are backed up only in their primary data center as they are not AHA capable.

Regular, automated tests are run to ensure the quality of backups, and any failures are reported for remediation within ServiceNow.

Transfer and failover

A transfer of an instance is a scheduled event that is usually performed for maintenance purposes, while a failover of an instance is an event usually performed where availability for one or more customer instances cannot be maintained.

ServiceNow has an established transfer process, but in the event that a failover is required, some of the process steps may be bypassed, as the active instance may not be accessible.

Critical system resources, including DNS, email, ServiceNow’s cloud operations systems, and customer service system are operated in high availability configurations in a minimum of two data centers.
In both the AHA transfer and failover scenarios, the cloud automation platform will make the customer instance in the secondary data center active. For further details on transfers and failovers, see the Transfer & Failover section in Securing the Now Platform.

Conclusion

ServiceNow is committed to providing cloud services that are always highly available, with built-in redundancy across all network and server infrastructure. All customer instances are individually provisioned on advanced multi-instance architecture that ensures that there can be no comingling of customer data. Data centers are arranged in pairs providing near-instant transfer from one data center to another providing Advanced High Availability for all customer production instances. Upgrades can be performed at a schedule determined by each individual customer, with no downtime required. Full backups are performed every seven days, with differential backups taken every 24 hours.

Finally, the ServiceNow cloud can easily be scaled horizontally to meet the needs of even the largest global enterprise. For further details on how ServiceNow delivers secure, scalable, and compliant cloud services, visit the ServiceNow/Trust site.