Advanced High Availability Architecture
Delivering performance, scalability, and availability with the Now Platform
Release: Washington, DC
Introduction

This document provides an overview of the ServiceNow® Advanced High Availability (AHA) architecture — a key element in delivering an enterprise-grade cloud service.

Organizations rely on access to IT and business data and services for their continued operation and success. The ServiceNow multi-instance architecture meets and exceeds stringent requirements surrounding data sovereignty, availability, and performance.

For an overview of the ServiceNow security program, please refer to the Securing the Now Platform white paper.

Please note: all information in this white paper is related to the standard Now Platform commercial environment.

For information related to other globally located ServiceNow in-country cloud offerings and how these offerings may differ, please contact your ServiceNow account representative.
Table of contents

Introduction ............................................. 2
Advanced High Availability overview ............... 4
Multi-instance architecture ........................... 5
Global data center pairs ............................... 5
Performance and scalability ......................... 6
Critical resources ..................................... 6
Data backup and recovery ............................ 7
Transfer and failover .................................. 7
   Transfer ............................................. 7
   Failover .......................................... 7
Conclusion ............................................. 8
Advanced High Availability overview

Please note: All information in this white paper is related to the standard Now Platform® commercial environment.

ServiceNow data centers and cloud-based infrastructure is 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 redundant power sources.

ServiceNow data centers are arranged in pairs, with all customer production data hosted in both data centers simultaneously and kept in sync using asynchronous database replication.

Both data centers are always active in a main-main relationship, with data replicated from the active (read-write) data center to the passive (read-only) data center.

Within each regional data center pair, there is no concept of a fixed primary location; each single data center (in a pair) is implemented to support the combined production load of both locations.

ServiceNow leverages Advanced High Availability (AHA) architecture 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 unlikely 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.

Instances of the Now Platform are deployed on an advanced, multi-instance architecture that provides separate application nodes and database processes for each customer.
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 customer’s instances. Customers can therefore 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

The ServiceNow data centers are arranged in regional pairs spanning five continents: Asia, Australia, Europe, North America, and South America.

ServiceNow uses top-tier global data center providers. These providers have no logical access to any ServiceNow systems or customer data and solely provide private colocation spaces and environmental resources. Only ServiceNow personnel with direct responsibility for (or role in) maintaining colocation spaces can physically access data center locations.

ServiceNow also offers data center pairs exclusively for qualified US Federal and Swiss banking customers. Meeting regulatory and sovereignty obligations is a significant factor when selecting data center facilities within specific geographic boundaries.
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. The ServiceNow Site Reliability Engineering (SRE) center employs a follow-the-sun model that provides continual 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 cloud operations systems, and customer service systems are operated in high availability configurations in a minimum of two data centers. None of these resources rely upon the internal ServiceNow 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.

Because ServiceNow uses the same AHA transfer process for scheduled preventative maintenance, the need for an annual disaster recovery test is eliminated.
Data backup and recovery

The ServiceNow Advanced High Availability (AHA) architecture is the primary means to restore service in the case of a disruption that could impact availability.

However, in certain scenarios, it may be desirable to use more traditional data backup and recovery mechanisms. For example, if a customer deletes some data inadvertently, or a customer’s data integration or automation gets misconfigured or malfunctions, this could result in data being rendered unusable or inaccessible. In these scenarios, the high availability architecture would not be applicable, and restoring from backup is the only option for recovery.

Full direct-to-disk backups are performed every seven days, and retained for a minimum of 14 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.

All backups are written to disk; tapes and removable media are not used. Backups are not sent off-site, but they are made within both data centers in a pair, therefore benefiting from geographic separation.

Backups are encrypted with AES-256 using randomly generated encryption keys for every backup. These are kept in a secure key store. And it is only retrieved by an automated process if a data restore is initiated. Regular, automated tests are run to ensure the quality of backups, and any failures are reported for remediation within ServiceNow.

The ServiceNow backup architecture is not designed to provide archival records, given the maximum 14-day default backup retention period. However, customers may retain data within their instances for as long as they require in accordance with their policy or regulatory requirements. Additionally, there are capabilities available within the Now Platform to allow customers to manage their unique instances logs and regularly export data to external systems, as required.

Transfer and failover

ServiceNow has two distinct processes related to ensuring instance availability: transfer and failover.

Transfer

A transfer of an instance is a scheduled event, usually performed for maintenance purposes and always coordinated with a customer. These outages occur within the contracted availability service level agreement that ServiceNow commits to with its customers.

Failover

A failover of an instance is an event usually performed where availability for one or more customer instances cannot be maintained. This could be down to a local component failure, or an event such as a major environmental incident or resource outage.

In the case of a local component failure, a failover to a system within the same data center will be attempted first. Where a data center–wide outage is identified, all current active production instances in the impacted data center will be failed over to the passive data center location in the pair. In this circumstance, a recovery time objective (RTO) of two hours, and a recovery point objective (RPO) of one hour is targeted.

Due to the almost real-time replication between data centers, these times are usually significantly shorter than the stated RTO/RPO.
Automation technology built on the Now Platform is used to transfer or failover instances when necessary. The mechanism for both processes is very similar. The current passive system is designated active, and vice versa. To complete the process, DNS mappings and instance database configurations are updated accordingly. Redundant DNS providers and DNSSEC are employed to provide robust, resilient name resolution services.

**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 which ensures that there can be no commingling 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](https://www.servicenow.com/trust).

> All customer instances are individually provisioned on advanced multi-instance architecture which ensures that there can be no commingling of customer data.