DELIVERING PERFORMANCE, SCALABILITY, AND AVAILABILITY
ON THE SERVICENOW NONSTOP CLOUD
Overview

Organizations, regardless of size, rely upon access to IT and business data and services for their continued operation and success.

This eBook provides an overview of the ServiceNow Advanced High Availability (AHA) architecture—a key element in delivering a true enterprise-grade cloud, the ServiceNow Nonstop Cloud. The unique, multi-instance architecture meets and exceeds stringent requirements surrounding data sovereignty, availability, and performance.

Advanced High Availability Architecture

ServiceNow’s data centers and cloud-based infrastructure are designed to be highly available. All servers and network devices have 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. Production application servers are load balanced within each data center. Production database servers are replicated in near-real time to a peer data center within the same geographic region in Asia, Australia, Europe, North America, and South America.

We leverage AHA for customer production instances in several ways:

- In the event of the failure of one or more infrastructure components, service is restored by transferring the operation of customer instances associated with the failed components to the peer data center.

- Before executing required maintenance, ServiceNow can proactively transfer operation of customer instances impacted by the maintenance to the peer data center. The maintenance can then proceed without impacting service availability.

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.
Global Data Center Pairs

ServiceNow’s data centers are arranged in pairs. There are six support centers and eight data center pairs for a total of 16 data centers. Our data center pairs span 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. A production instance for one customer may be primary out of one data center in the pair while a production instance for another customer may be primary from the other.

ServiceNow maintains continuous, asynchronous replication from the database in the current primary data center (read-write) to the secondary data center (read-only). To transfer a customer instance from a primary to a secondary data center, ServiceNow designates the secondary to be the primary and the primary to be the secondary.
Multi-Instance Architecture

The ServiceNow Nonstop Cloud is deployed on an advanced, multi-instance architecture that provides separate application nodes and database processes for each customer. This means there is no co-mingling of customer data in a shared database that would exist in a multi-tenant architecture.

We deploy instances on a per-customer basis, allowing the multi-instance cloud to scale horizontally to meet each customer’s performance needs.

Unlike in a multi-tenant environment, each instance runs its own application logic and database processes. This means your instance does not have to be on the same version or upgraded at the same time as other customers’ instances. You can choose to upgrade your instances on a schedule that best meets your enterprise’s needs and compliance requirements.
Availability

The ServiceNow Nonstop Cloud aims to be always operational for our customers. No vacation, no extended upgrade or maintenance windows, no single points of failure. We focus on near-perfect availability with redundancy built into every layer of our cloud, including redundant devices and power across all network and server infrastructure. There is no downtime necessary for upgrades.

We provide the industry’s only Real Availability Dashboard that shows availability of all your instances running in the cloud. Real Availability is the true measure of customer availability by looking at every incident that results in a customer outage (a Priority 1 or P1 incident).

Performance

Our Nonstop Cloud scales to meet the needs of the largest Global 2000 enterprises and aims to be always operational for our customers. We have tens of thousands of customer instances operating globally in our data center regions. Each of our customer instances leverages our multi-instance architecture to perform an aggregate of tens of billions of full page transactions every month. Customers using the ServiceNow Configuration Management Database (CMDB) as the single system of record have scaled their CMDBs to manage tens of millions of configuration items (CIs).
AHA Process Overview

The AHA process is comprised of eight main steps and it is invoked through using an instance of Now Platform (using an implementation of ServiceNow to orchestrate the Nonstop Cloud) when one of two conditions is met:

1. In the event of a service disruption, the ServiceNow operations team determines whether a failover is required.
2. For scheduled maintenance activity, the ServiceNow operations team determines if an AHA transfer should be performed.

The main automated AHA transfer steps are as follows:

1. Run an end-to-end automated suite of pre-flight checks to ensure that all infrastructure and application configurations associated with the customer’s active and standby instances are in a healthy state, including data replication between the data centers. During the pre-flight checks, the customer instance is fully available and operational.
2. Change the Domain Name Service (DNS) information associated with the customer instance.
3. Stop all application nodes associated with the customer instance.
4. Reverse the roles for each database from active (read-write) to passive (read-only) and vice versa.
5. Change the database pointer to the read-write database within the application nodes.
6. Re-start all application nodes associated with the instance. At this point, the instance is again fully available.
7. Run an end-to-end automated suite of post-flight checks to ensure all systems and configurations are in a healthy state.
8. Perform discovery so that the configuration management database (CMDB) is updated with the new configuration.

In the event an AHA failover is required, some of the above steps are bypassed, as the active instance may not be accessible. In both the AHA transfer and AHA failover scenario, the cloud automation platform will make the customer instance in the secondary data center active.
Backup and Recovery

While Advanced High Availability is the primary means to recover data and restore service in the case of a disruption, in certain cases it is desirable to use ServiceNow’s more traditional data backup and recovery mechanism. It works in concert with AHA and acts as a secondary recovery mechanism.

Backups of the two production databases and the single sub-production database are taken every day for all instances.

The backup cycle consists of four weekly full backups and the past six days of daily differential backups that provide 28 days of backups. All backups are written to disk, no tapes are used and no backups are sent off site. All the controls that apply to live customer data also apply to backups. If data is encrypted in the live database, then it will also be encrypted in the backups.

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

Critical Resources

ServiceNow is responsible for managing its environment, the supporting infrastructure, and vendor relationships. As part of these responsibilities, ServiceNow maintains a 24x7 Site Reliability Engineering (SRE) Center to monitor uptime and availability. The SRE uses a “follow-the-sun” model, which provides continuous security, operational monitoring and support of the ServiceNow environment and infrastructure. ServiceNow rotates operations and technical support daily—in North America, the U.K., and Australia—in order to provide 24x7 operations and security monitoring.

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.

We use AHA for our 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. It lets them support and continue developing 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.

Summary

We believe that cloud services must always be on. Our unique, multi-instance architecture lets you configure your cloud services and perform upgrades on your own schedule. And our advanced, high availability infrastructure provides instance redundancy between data center clusters in your chosen geography. It scales to meet the needs of even the largest global enterprises.

For more information on how ServiceNow delivers secure, scalable, and compliant cloud services, visit www.servicenow.com/trust.

Delivering Performance, Scalability, and Availability