SoftNAS High Availability Deployment Guide

SoftNAS SNAP HA High Availability delivers a low-cost, low-complexity solution for high-availability clustering that is easy to deploy and manage. A robust set of HA capabilities protect against data center, availability zone, server, network and storage subsystem failures to keep business running without downtime. SNAP HA for Amazon Web Services (AWS) includes patent-pending Elastic Load Balancing technology, providing NAS clients in any availability zone uninterrupted HA access to the storage cluster across availability zones.

SNAP HA offers two methods of data protection, based on the storage type selected at creation, Standard or Shared Pools:

SNAP HA

Several measures have been taken to ensure the highest possible data integrity of your highly available block storage system. An independent "witness" HA controller function ensures there is never a condition that can result in what is known as "split-brain", where a controller with outdated data is accidentally brought online. SNAP HA prevents split-brain using a number of industry-standard best practices, including use of a 3rd party witness HA control function that tracks which node contains the latest data. On AWS, shared data stored in highly redundant S3 storage is used. On VMware, a separate HA Controller VM is used.

Another HA feature is "fencing". In the event of a node failure or takeover, the downed controller is shut down and fenced off, preventing it from participating in the cluster until any potential issues can be analyzed and corrected, at which point the controller can be admitted back into the cluster.

Finally, data synchronization integrity checks prevent accidental failover or manual takeover by a controller which contains data which is out of date.

The combination of high-integrity features built into SNAP HA ensures data is always protected and safe, even in the face of unexpected types of failures or user error.

Note: Even with these strong measures in place, limited data loss (approximately 5 seconds worth) can occur at the moment of failure if default settings for SoftNAS' implementation

are used. This risk is present to a varying degree in any high availability solution relying on the real-time transfer of active data between two nodes. SoftNAS' default settings are in place to provide a balance between performance and data integrity concerns. Measures can be taken when creating pools and volumes for high availability to limit or eliminate this potential loss. Sync mode settings can be used to further enforce data integrity, but with a hit to performance. SoftNAS strongly recommends the creation of a write log, or ZIL to cache high bursts of write activity, and further protect data integrity, as well as boosting performance.

Minimize Downtime from Host and Storage Failures

SoftNAS SNAP HA High Availability delivers the availability required by mission-critical applications running in virtual machines and cloud computing environments, independent of the operating system and application running on it. HA provides uniform, cost-effective failover protection against hardware and operating system outages within virtualized IT and cloud computing environments. SNAP HA:

In AWS, SNAP HA is applied to SoftNAS storage controllers running in a Virtual Private Cloud (VPC). It is recommended to place each controller into a separate AWS Availability Zone (AZ), which provides the highest degree of underlying hardware infrastructure redundancy and availability.

AWS Minimum Requirements

- Virtual Private Cloud (VPC)
- 1 Elastic or Virtual IP address, used to route NAS client traffic across availability zones
- 2 SoftNAS storage controller EC2 instances
- Amazon S3 storage (2 MB of S3 storage will be allocated in same region as EC2 instances.)
- Two 1 GB virtual interfaces on each instance
- EBS disks and/or S3 Cloud Disks for each storage controller's local storage

VMware Minimum Requirements

- 2 SoftNAS storage controller VMs
- HA Controller VM is required, with a recommended min. of 500 MB RAM and 1 vCPU
- One 1 Gb virtual NIC (shared for admin, replication and HA monitor - best practice is to use multiple NICs)
- Two 1 GbE physical NICs
- One or more VMDK virtual disks for storage

Azure Minimum Requirements

- Virtual Private Cloud (VPC)
- 2 SoftNAS storage controller Azure Virtual Machines
- Two 1 GB virtual interfaces on each instance
- Azure Standard/Premium block storage disks or Azure Hot or Cold Blob object storage.

Recommended Configurations

The following configurations are recommended best practices for SoftNAS SNAP HA:

- 16 to 64 GB RAM
- 4 vCPU (8 vCPU if volume data compression will be used extensively)
- SSD for read cache and write log
- Separate replication and storage traffic to dedicated physical networks
- Replication should occur from like storage to like storage, to avoid performance bottlenecks. Different read/write speeds can potentially result in cache data loss.

Note: SNAP HA relies on time settings in both the primary and secondary instances. It is important to use the same time for each. If an NTP is used, configure both with the same URL.

AWS Requirements

- Virtual Private Cloud (VPC)
- 3 Virtual IP addresses
- one used to route NAS client traffic across availability zone. This IP address must be in a separate CIDR block.
- one for each instance for SoftNAS StorageCenter remote administration. Virtual IP setup is recommended.
- Alternatively, 3 Elastic IP addresses:
- one used to route NAS client traffic across availability zones,
- one each for SoftNAS StorageCenter remote administration.
- Alternatively, use a VPC with private VPN access to SoftNAS StorageCenter for administration, with 1 Elastic IP address for NAS client traffic
- 2 each SoftNAS storage controller EC2 instances
- Amazon S3 storage (2 MB of S3 storage will be allocated in same region as EC2 instances
- 2 virtual interfaces on each instance. First interface is used for SoftNAS StorageCenter and replication, second interface for Elastic HA IP for NAS traffic
- For storage VLAN, choose EC2 instance types for NAS clients and SoftNAS StorageCenter that support MTU 9000 (required for 10 GbE maximum throughput)

- EBS disks and/or S3 Cloud Disks for each storage controller's local storage
- For highest throughput, use HVM instances with local, ephemeral SSDs for read cache, high-IOPS EBS volume (SSD) for write log and EC2 instances with 10 GbE network interfaces
- Use EBS volumes for primary storage in RAIDz-2 configuration for best data density and RAID-10 with high-IOPS EBS volumes for best IOPS in database and transactional applications
- Use S3 disks for lower IOPS, highly redundant mass-storage up to 4 PB per S3 disk device

VMware Requirements

- HA Controller VM is required, the recommended minimum is 500MB of RAM and 1vCPU
- 2 each SoftNAS storage controller VMs
- 1 each HA Controller VM with 500 MB RAM and 1 vCPU configured to use VMware FT (fault-tolerance) to ensure HA Controller is always available
- 3 each virtual NICs - separate vNIC and VLAN allocated to: 1) SoftNAS StorageCenter administration (E1000), 2) SnapReplicate block replication (E1000), 3) storage VLAN (VMXNet3)
- For storage VLAN, configure for MTU 9000 (required for 10 GbE maximum throughput)
- DirectPath pass-through disk controller providing direct disk access (requires Intel VT-d and disk controller supported by CentOS). This is required for best small block 4K/8K I/O and synchronous write-log and read cache performance with VMware
- Separate disk controllers for 1) booting VMware from RAID-1 mirrored disks and 2) storage I/O
- 4 each 10 GbE or 1 GbE physical NICS (2 active/active for VMware host management and SoftNAS administration and replication, 2 active/active for data storage)

Optional

- Boot VMware from 32 GB USB, and dedicate disk controller for DirectPath disk I/O
- VMDKs for SATA and SAS storage and read cache
- Infiniband NIC for data storage pathway

Operation in AWS Virtual Private Cloud

In AWS, SNAP HA is applied to SoftNAS storage controllers running in a Virtual Private Cloud (VPC). It is recommended to place each controller into a separate AWS Availability Zone (AZ), which provides the highest degree of underlying hardware infrastructure redundancy and availability.

Virtual IP Setup

Each AZ operates on a separate subnet; e.g., -10.0.1.0/24 and 10.1.0.0/24 (choose how to organize the subnet addresses in the VPC based on expected requirements). SoftNAS SNAP HA can now take advantage of Virtual IPs. One virtual IP address is assigned to each VPC instance, set up within the same CIDR block. A third lone IP address is set up on a separate CIDR block, to manage NAS client traffic requirements.

Virtual IPs are isolated from internet traffic completely, increasing the security of your HA VPC setup. For this reason, a Virtual IP driven private HA setup is our recommended best practice.

HA storage traffic uses a dedicated network interface (interface 1), which further isolates storage traffic.

Elastic IP setup

Traditionally, an elastic IP provided NAS clients across all AZs access to HA storage. Until recently Elastic IPs were the only IPs capable of re-routing network traffic across AZs. SoftNAS SNAP HA enhances the standard elastic IP provided by AWS, creating a patent-pending "Elastic HA" (EIP). Elastic HA IPs are managed by the HA controller, ensuring NAS client traffic is properly routed to the active primary storage controller at all times.

HA storage traffic uses a dedicated network interface (interface 1), which further isolates storage traffic.

There's a common misconception that elastic IPs are only useful for Internet-based access to EC2 instances. While that is the most common use case by far, Elastic HA IP addresses are typically configured using a Security Group which restricts access within the AZ private network only. This prevents any possible Internet-based access to Elastic HA IPs.

VPCs can also be configured for use with VPNs, which enables secure access from an administrator's office location to the private network (no other inbound Internet access is typically available). It is possible to attach optional elastic IP addresses to interface 0 on each SoftNAS controller instance for remote administration (restricted IP range access recommended).

Operation in VMware Private Clouds

On VMware, it is common to dedicate a non-routable VLAN to storage traffic. The storage VLAN segregates primary storage traffic (e.g., VMDKs attached to VMs over NFS or iSCSI) from other traffic. Data replication traffic can also be placed on its own separate non-

routed VLAN. SoftNAS StorageCenter is typically placed on a routable VLAN (the default network), where it can be readily accessed by admins from a web browser from anywhere within the organization (or via a VPN).

A Virtual IP (VIP) address is employed to route NAS client traffic to the primary storage controller. In the event of a failover or takeover, the VIP is reassigned to the other controller, which immediately re-routes NAS client traffic to the proper controller.

Operations in Azure - Availability Sets

Availability Sets make use of two key concepts - Fault Domains, and Update Domains. At its core, Azure consists of racks upon racks of servers. Each rack can host any number of virtual machines. When creating a highly available pairing, you want to be sure that there is no single point of failure, that your workload will still be provisioned by one virtual machine if the other is under maintenance. Unfortunately, if you do not specify otherwise, there is no guarantee that your VMs will not be placed on the same rack, or the same 'Fault Domain'. In essence, a fault domain can be considered a rack within Azure. Every VM on the rack is subject to that rack's power and network connections. A rackwide failure, or a rackwide maintenance window will take down all VMs hosted on this single point of failure. When Azure refers to a fault domain, consider each fault domain a single point of failure.

An Availability Set distributes highly available workloads across multiple Fault Domains, thereby eliminating any single point of failure. Unless the entire data center is down, your workload will keep running. In essence, your workload is split between two or more racks, leveraging the redundant power supplies, network switches, etc, of each.

data.