

NUTANIX TECH NOTE

Nutanix Clusters on AWS

Copyright

Copyright 2020 Nutanix, Inc.
Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws.

Nutanix is a trademark of Nutanix, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

| | |
|---|----|
| 1. EXECUTIVE SUMMARY | 5 |
| 2. NUTANIX CLUSTERS PORTAL | 6 |
| 3. NUTANIX CLOUD NETWORKING | 9 |
| 3.1. Creating a Subnet | 11 |
| 4. MIGRATION | 18 |
| 5. STORAGE AVAILABILITY IN AWS | 19 |
| 5.1. Deal with Failures | 22 |
| 5.2. Prevent Network Partition Errors | 23 |
| 5.3. Proactively Resolve Bad Disk Resources | 23 |
| 5.4. Maintain Availability: Disk Failure | 24 |
| 5.5. Maintain Availability: Availability Zone Failure | 24 |
| 6. DEPLOYMENT MODELS | 25 |
| 6.1. Multicloud Deployment | 26 |
| 6.2. Multiple-Availability Zone Deployment | 27 |
| 6.3. Single-Availability Zone Deployment | 29 |
| 7. CAPACITY OPTIMIZATION | 36 |
| 7.1. Compression | 36 |
| 7.2. Deduplication | 36 |
| 8. ENCRYPTION | 38 |
| 9. VIRTUAL MACHINE HIGH AVAILABILITY | 39 |
| 9.1. VMHA Recommendations | 40 |
| 10. ACROPOLIS DYNAMIC SCHEDULER | 41 |
| 10.1. Affinity Policies | 41 |
| 11. CONCLUSION | 42 |
| APPENDIX | 43 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1: Overview of the Nutanix Hybrid Multicloud Software | 5 |
| Figure 2: Clusters Management | 7 |
| Figure 3: Clusters Portal | 7 |
| Figure 4: OVS Conceptual Architecture | 10 |
| Figure 5: IPAM with AWS | 12 |
| Figure 6: AWS Networking | 13 |
| Figure 7: Cluster Deployment Network on AWS | 14 |
| Figure 8: Cluster Deployment Configuration on AWS | 15 |
| Figure 9: VPN Connection | 16 |
| Figure 10: AWS Security Groups | 17 |
| Figure 11: Partition Placement | 19 |
| Figure 12: Mixing and Matching | 20 |
| Figure 13: Replication Factor Data Placement Across Racks | 21 |
| Figure 14: Data Path Redundancy | 23 |

| | |
|--|----|
| Figure 15: Multicloud Deployment..... | 25 |
| Figure 16: Cluster Details..... | 27 |
| Figure 17: Multiple-Availability Zone Nutanix Clusters Deployment..... | 28 |
| Figure 18: Private Route Table..... | 29 |
| Figure 19: HYCU Backup Architecture..... | 30 |
| Figure 20: S3 VPC Endpoint..... | 31 |
| Figure 21: Block Public Access to Backup S3 Bucket..... | 31 |
| Figure 22: Enable Versioning to Use CRR..... | 32 |
| Figure 23: Veeam Backup Architecture..... | 32 |
| Figure 24: Backup Configuration of AHV Proxy..... | 34 |
| Figure 25: Backup Copy Job..... | 35 |
| Figure 26: VM High Availability Reservation..... | 39 |

LIST OF TABLES

| | |
|--|----|
| Table 1: Document Version History..... | 6 |
| Table 2: Desired Fault Tolerance and Required Nodes..... | 21 |
| Table 3: Recommendations for Replication Factor..... | 22 |
| Table 4: Cluster Outbound to the Cluster Portal..... | 25 |
| Table 5: Cluster Outbound to EC2..... | 25 |
| Table 6: Cluster Outbound to Python Package Index..... | 25 |
| Table 7: Inbound Security Group Rules for AWS..... | 26 |
| Table 8: Availability Zone 1 NCA Security Group settings..... | 28 |
| Table 9: Availability Zone 2 NCA Security Group Settings..... | 28 |
| Table 10: Veeam Ports between the Nutanix Cluster and VBR..... | 33 |
| Table 11: Nutanix Compression Policies..... | 36 |
| Table 12: Recommendations for Deduplication per Container (Datastore)..... | 37 |



1. Executive Summary

Nutanix designed its software for customers running their workloads on cloud computing providers like Amazon Web Services (AWS) to have the same experience they're used to with on-premises Nutanix clusters. Because Nutanix Clusters on AWS (NCA) runs Nutanix AOS and AHV with the same CLI, UI, and APIs, existing IT processes or third-party integrations that work on-premises continue to work regardless of where they're running.

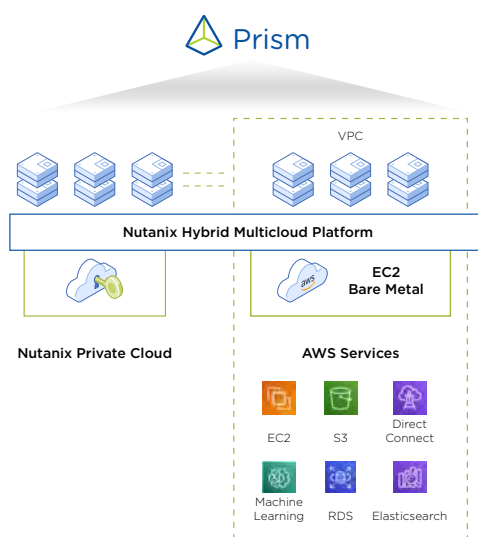


Figure 1: Overview of the Nutanix Hybrid Multicloud Software

NCA situates the complete Nutanix hyperconverged infrastructure (HCI) stack directly on an Amazon Elastic Compute Cloud (EC2) bare-metal instance. This bare-metal instance runs a Controller VM (CVM) and Nutanix AHV as the hypervisor just like any on-premises Nutanix deployment, using the AWS elastic network interface (ENI) to connect to the network. AHV user VMs do not require any additional configuration to access AWS services or other EC2 instances.

AHV runs an efficient embedded distributed network controller that integrates user VM networking with AWS networking. AHV assigns all user VM IPs to the bare-metal host where VMs are running. Instead of creating an overlay network, the AHV embedded network controller simply provides the networking information of the VMs running on NCA, even as a VM moves around the AHV hosts. Because NCA integrates IP address management with AWS Virtual Private Cloud (VPC), AWS allocates all user VM IPs from the AWS subnets in the existing VPCs.

AOS can withstand hardware failures and software glitches and ensures that application availability and performance are never compromised. Combining features like native rack awareness with AWS partition placement groups allows Nutanix to operate freely in a dynamic cloud environment.

NCA quickly gives on-prem workloads a home in the cloud, offering native access to available cloud services without requiring you to reconfigure your software.

Table 1: Document Version History

| Version Number | Published | Notes |
|----------------|-------------|----------------------|
| 1.0 | August 2020 | Original publication |



2. Nutanix Clusters Portal

Customers access the Clusters portal through their existing accounts at my.nutanix.com. You can use the portal to deploy AWS clusters and to manage tasks like health remediation and expanding and condensing your clusters. On-prem Prism Central can manage your deployed NCA alongside your on-prem clusters. For easy day-two operations, Prism Central can also manage AOS upgrades for on-prem, remote or branch office, and cloud-based Nutanix clusters.

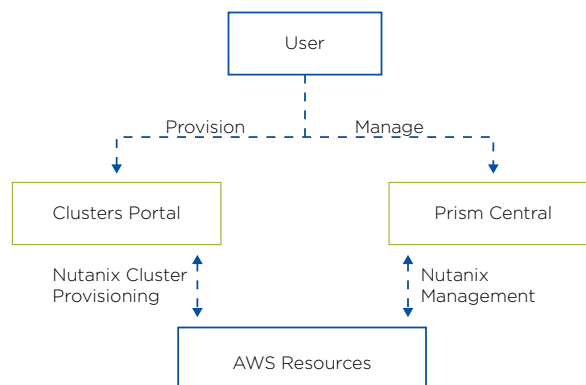


Figure 2: Clusters Management

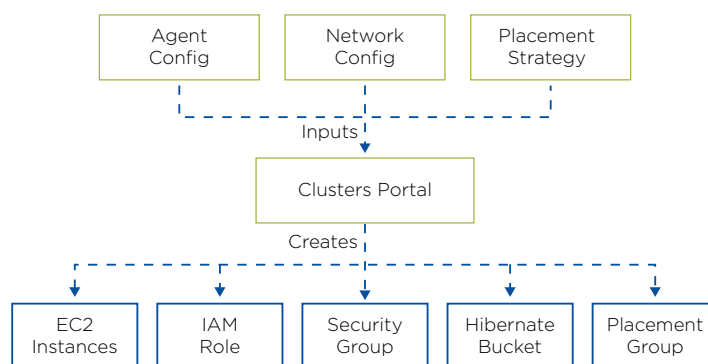


Figure 3: Clusters Portal

The Clusters Portal provides the following services:

- Obtaining and managing bare-metal resources.
- Ensuring the correct IAM roles are created and used for deployment.
- Creating AWS security group rules to help lock down your AWS resources.
- Performing hibernate and resume operations, including creating S3 buckets.
- Managing node placement strategy and removing or adding nodes based on the health of the cluster.



3. Nutanix Cloud Networking

Nutanix can deliver a true hybrid multicloud experience because it has native cloud networking. Nutanix integration with the AWS networking stack means that every VM deployed on an NCA cluster gets a native AWS IP, so as soon as you migrate or create an application on NCA, it has full access to all AWS resources. It also removes the burden of managing and deploying an additional network overlay. Since the Nutanix network capabilities are directly on top of the AWS overlay, network performance remains high and host resources aren't consumed by additional network controllers.

Because it has native network integration, you can deploy NCA in existing AWS VPCs. Since existing AWS environments have gone through change control and security processes already, you don't need to do anything except allow the NCA to talk to a Clusters Portal. We believe that doing so enables our customers to increase security in their cloud environments.

Nutanix uses native AWS API calls to deploy AOS on bare-metal EC2 instances and consume network resources. Each bare-metal EC2 instance has full access to its bandwidth through an elastic network interface (ENI). For example, if you deploy Nutanix to an i3.metal instance, each node has access to 25 Gbps. With AHV, the ENI ensures that you don't need to set up additional networking high availability for redundant network paths to the top-of-rack switch.

AHV uses Open vSwitch (OVS) for all VM networking. You can configure VM networking through Prism or the aCLI, and each vNIC connects to a tap interface. The following figure shows a conceptual diagram of the OVS architecture.

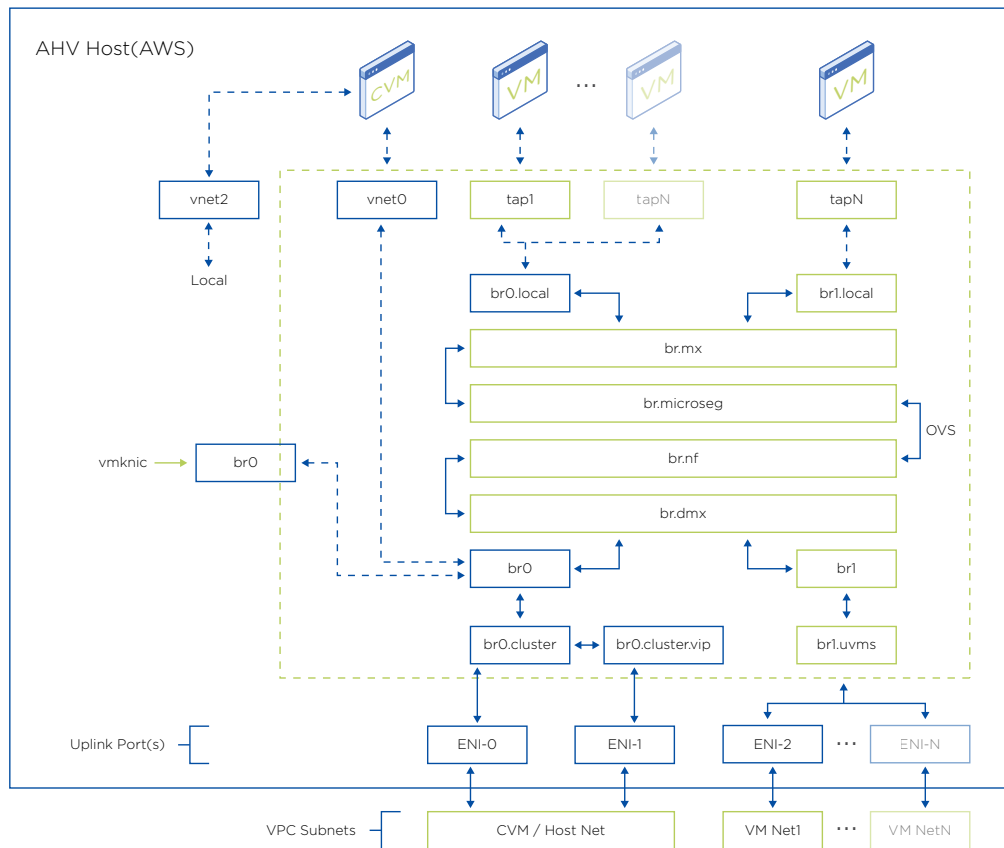


Figure 4: OVS Conceptual Architecture

When AOS runs on AWS, you can rely on the AWS overlay network to provide the best possible throughput automatically, leading to a very consistent and simple network configuration.

The ENI is a logical networking component in a VPC that represents a virtual network card. An ENI can have one primary IP and up to 50 secondary IPs. All deployed user VMs (UVMs) use the secondary IPs to get direct AWS network access. AHV hosts deployed in AWS have separate

ENIs for management traffic (AHV/CVM) and UVMs, which means customers can have different AWS security groups for management and UVMS. NCA creates a single default security group for UVMs running in the Nutanix cluster. Any ENIs created to support UVMs are members of this default security group, which allows all UVMs in a cluster to communicate with each other. In addition to the security groups, Nutanix customers can use the microsegmentation built into Nutanix Flow to provide greater security controls for east-west network traffic.

3.1. CREATING A SUBNET

A customer first creates a subnet in AWS in a VPC (virtual private cloud), then connects it to AOS in Prism Element. Cloud Network, a new service in the CVM, works in conjunction with AOS configuration and assigns a VLAN ID (or VLAN tag) to the AWS subnet and fetches relevant details about the subnet from AWS. The network service keeps customers from using the AHV or CVM subnet for UVMs by not allowing them to create a network with the same subnet.

You can use each ENI to manage 49 secondary IP addresses. A new ENI is also created for each subnet you use. The AHV host, VMs, and physical interfaces use ports to connect to the bridges and both bridges communicate with the AWS overlay network. Because each host already has the drivers needed for a successful deployment, you don't need to do any additional work to use the AWS overlay network. Just keep the following best practices in mind:

- Don't share AWS user VMs subnets between NCA instances.
- Have separate subnets for Management (AHV/CVMs) and user VMs.
- If you plan to use VPC peering, use nondefault subnets to ensure uniqueness across AWS Regions.
- Divide your VPC network range evenly across all usable Availability Zones in a Region.
- In each Availability Zone, create one subnet for each group of hosts that has unique routing requirements (for example, public versus private routing).
- Size your VPC CIDR and subnets to support significant growth.

AHV IP Address Management (IPAM)

AHV uses IP address management (IPAM) to integrate with native AWS networking. NCA uses the native AHV IPAM to inform the AWS DHCP server of all IP assignments using API calls. NCA relies on AWS to send gratuitous ARP packets for any additions to an ENI's secondary IP addresses. We rely on these packets to ensure that each hypervisor host is notified when an IP address moves or new IP addresses become reachable. A given AWS subnet can't be shared between two NCAs.

We added a new service called the cloud network controller (CNC) to the AHV host to help with ENI creation. CNC runs an OpenFlow controller, which manages the OVS in the AHV hosts and handles mapping, unmapping, and migrating UVM secondary IP addresses between ENIs or hosts. A subcomponent of CNC called cloud port manager provides the interface and manages AWS ENIs.

Update Network

VLAN ID

0

☒ ENABLE IP ADDRESS MANAGEMENT

NEW!

This gives Acropolis control of IP address assignments within the network.

NETWORK IP ADDRESS / PREFIX LENGTH

192.168.101.0/24

GATEWAY IP ADDRESS

192.168.101.1

☐ CONFIGURE DOMAIN SETTINGS

IP ADDRESS POOLS

+ Create Pool

| START ADDRESS | END ADDRESS | |
|----------------|-----------------|-----|
| 192.168.101.20 | 192.168.101.200 | ✎ ✕ |

Figure 5: IPAM with AWS

IPAM avoids address overlap by sending AWS API calls to inform AWS which addresses are being used.

The Acropolis master assigns an IP address from the address pool when it creates a managed vNIC and releases the address back to the pool when the vNIC or VM is deleted.

NOTE: You can't use or assign the first four IP addresses or the last IP address in each subnet.

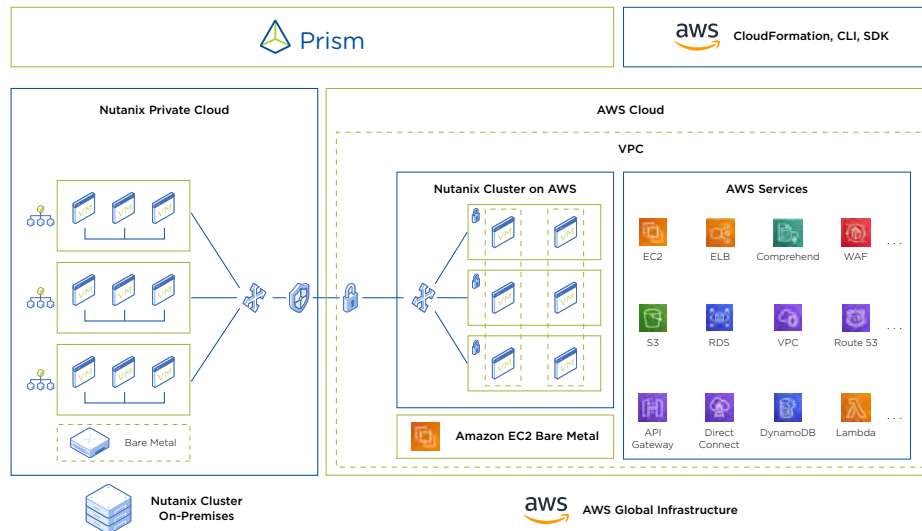


Figure 6: AWS Networking

Using native AWS networking allows you to quickly establish connectivity and eliminate costly performance impacts from third-party networking overlays. Cloud administrators can focus on their tasks instead of managing additional networking technologies. Let's walk through a typical deployment to see the AWS constructs in action.

- Click **Create Cluster** in the Cluster Portal.
- Provide the name and URL name for the cluster.
- Select **AWS** as the cloud provider.
- Fill in the other information for your specific cluster, then click **Validate Network Configuration** to test your configuration.
- Click **Next**.

Create Cluster

1 Region and Network 2 Configuration 3 Summary

Name: URL name:

Cloud provider:

Cloud Account:

Region: [Check availability](#)

Virtual Private Network (VPC): [Create new](#)

Subnet:

Availability zone: us-east-1c

Figure 7: Cluster Deployment Network on AWS

- In the Configuration section, select your AOS version.
- Configure the Cluster Capacity as desired, and make sure you select **Disabled** for Prism access from the public Internet.
- We selected **Terminate at a specific time** for Scheduled Account Termination because we are using this instance as a test. Use the setting that best suits your cluster needs.

Create Cluster

1 Region and Network
2 Configuration
3 Summary

AOS Version

5.11.2
⌵

Cluster Capacity

Replication Factor

–

2

+

Host type

i3.metal
⌵

Number of nodes

–

16

+
⛔

[Add Hosts](#)

Select SSH key [Create new](#)

nutanixdemo
⌵

Prism access from the public Internet

☐ Allowed
☒ Disabled
☐ Restricted

Scheduled account termination

☐ Do not terminate
☒ Terminate at specific time

Terminate on

Mar 23rd 2020 18:24
📅

Time zone

America/Denver (MDT)
GMT-06:00 ⌵

Figure 8: Cluster Deployment Configuration on AWS

After you fill in both parts of the configuration, click Submit. 30 to 40 minutes later your cluster is provisioned. Native integration with the AWS network provides flexibility and makes it easy to integrate with the other AWS services.

You can also use an existing AWS subnet that follows your architectural standards or create a new one. The subnet you select or create becomes the management subnet and provides the IP addresses for the hypervisor (Nutanix AHV) host and the CVM.

Once you deploy the cluster, you can set up a VPN gateway in AWS and create a site-to-site VPN connection. The following figure shows a high-level overview of a VPN connection for a typical NCA deployment.

Security Groups (1/3) [Info](#)

Filter security groups

search: 116 X Clear filters

| | Name | Security group ID | Security group name |
|-------------------------------------|------|----------------------|---------------------------------------|
| <input checked="" type="checkbox"/> | - | sg-0615b4e4d6051de41 | prod:cluster:1116:uvm |
| <input type="checkbox"/> | - | sg-06e36e22dd650f085 | prod:cluster:1116:internal_management |
| <input type="checkbox"/> | - | sg-07af307eb16ff4d45 | prod:cluster:1116:user_management |

Figure 10: AWS Security Groups

With AWS security groups, you can choose to allow access to the AWS CVMs and AHV host only from your on-prem management network and CVMs. You can lock down replication from on-prem to AWS at the granularity of the specific port. Because all the replication software is embedded in the CVM on both sides, you can easily migrate your workloads back and forth.

The simplicity of Nutanix Clusters also saves you money. Because you don't need any additional overlay networks, you save on the cost for the additional compute the overlays require. You avoid the costs for management gateways, network controllers, edge devices, and storage incurred from adding appliances. With a simpler system, you also realize significant operational savings on maintenance and troubleshooting.



4. Migration

There are many reasons to move your applications to AWS, including consolidation, bursting, or wanting them near a cloud-based service. Once you configure networking from AWS to on-prem, you can choose any proven methods for moving applications to an AHV-based cluster, which saves time and money. The following are most common ways to migrate data to NCA:

- [Native data protection](#)

You can use this method for ESXi- and AHV-based clusters. Creating a remote site for your new NCA and setting up the native networking integration only takes a few minutes; you simply need to ensure that the ports are open on the management security group you need for replication. All the existing [data protection best practices](#) apply because a bare-metal AWS deployment essentially acts as an additional supported OEM vendor.

- [Leap \(DR orchestration\)](#)

If you want to take advantage of protection policies and recovery plans to protect applications across multiple Nutanix clusters, set up Leap from Prism Central by selecting the check box. Whether you're doing DR or migrations, Leap stages your applications to be restored in the right order. You can also use the protection policies to quickly revert back to on-prem if desired.

- [Nutanix Move](#)

Nutanix Move is a cross-hypervisor migration solution that migrates VMs with minimal downtime. Nutanix Move supports three migration types: VMs running on ESXi managed by vCenter, EBS-backed EC2 instances running on AWS, and VMs running on Hyper-V. Nutanix Move also supports migrating AWS EC2 VMs to AHV on the Nutanix cluster, though this use case is minimal.

- [AHV-based backups](#)

You can use any third-party backup product to restore applications to NCA, which is important when you need to migrate or do testing and development work.



5. Storage Availability in AWS

Nutanix uses a partition placement strategy when deploying nodes inside an AWS Availability Zone. An Availability Zone is a logical datacenter available for any AWS customer in that Region to use. An AWS Region is a separate geographic area. Each Region has multiple, isolated locations known as Availability Zones. Each zone in a Region has redundant and separate power, networking, and connectivity to reduce the likelihood of two zones failing simultaneously. One Nutanix cluster can't span different Availability Zones in the same Region but you can have multiple Nutanix clusters replicating between each other in different zones or Regions.

Using up to seven partitions, Nutanix places the AWS bare-metal nodes in different AWS racks and stripes new hosts across the partitions.

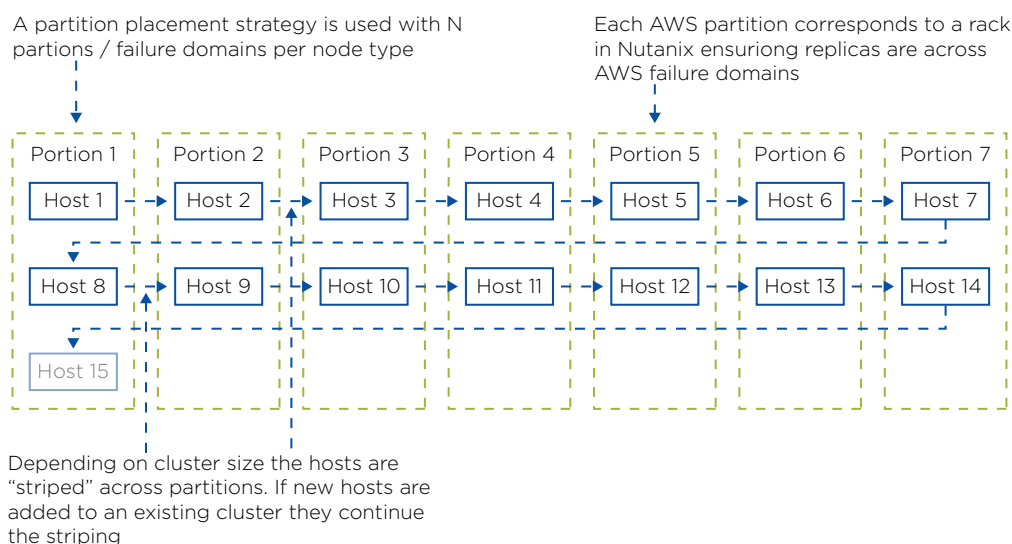


Figure 11: Partition Placement

The following diagram illustrates that you can add different node types to the same cluster. Because different node types go in different racks, the placement algorithm works the same way.

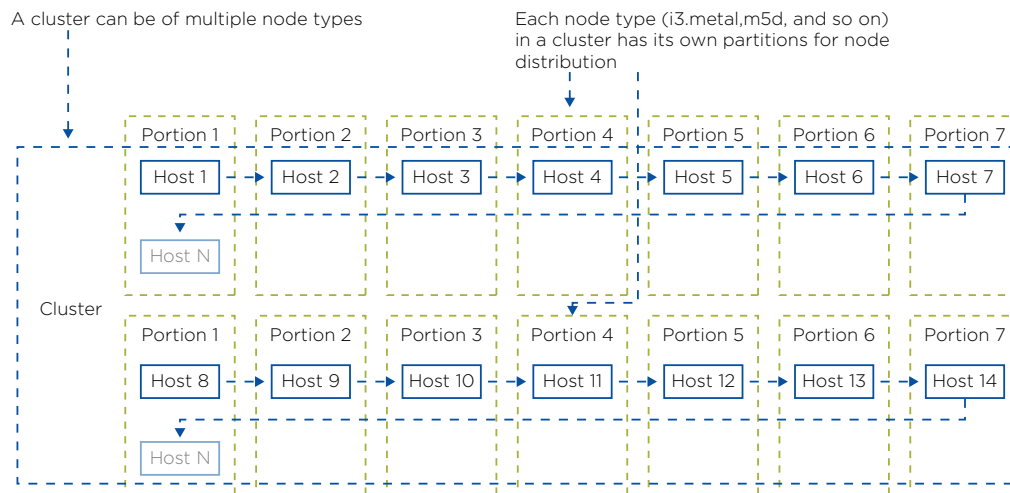


Figure 12: Mixing and Matching

When you have formed the Nutanix cluster, the partition groups map to the Nutanix rack-awareness feature. The DSF writes data replicas to other racks in the cluster to ensure that the data remains available for both replication factor 2 and replication factor 3 scenarios in the case of a rack failure or planned downtime.

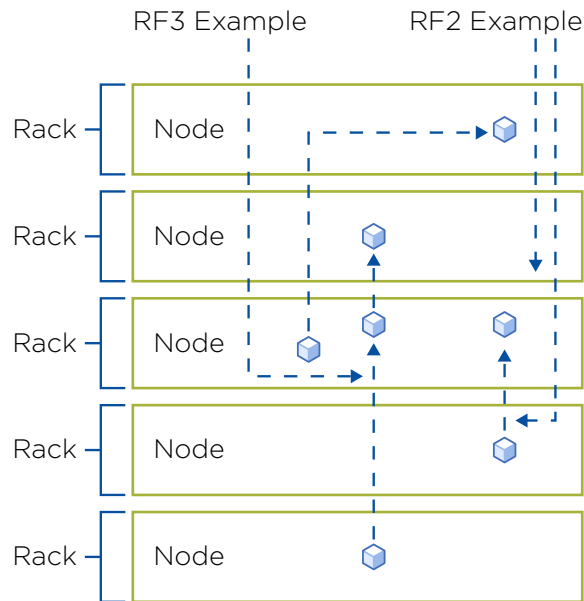


Figure 13: Replication Factor Data Placement Across Racks

The following table highlights the minimum number of racks required in your cluster to withstand a given number of rack failures. Nutanix Erasure Coding (EC-X) is one of the storage reduction technologies available in AOS. EC-X takes one or two data copies and creates a parity you can use to recreate the data if required.

Table 2: Desired Fault Tolerance and Required Nodes

| Desired Awareness Type | Fault Tolerance Level | EC-X Enabled | Minimum Units in the Cluster | Simultaneous Failure Tolerance |
|------------------------|-----------------------|--------------|------------------------------|--------------------------------|
| Rack | 1 | No | 3 Rack | 1 Rack |
| Rack | 1 | Yes | 4 Rack | 1 Rack |
| Rack | 2 | No | 5 Rack | 2 Rack |
| Rack | 2 | Yes | 6 Rack | 2 Rack |

Administrators can use replication factor 3 when high availability requirements exceed the data protection level that replication factor 2 provides. We also recommend using replication factor 3 in larger clusters (32 or more nodes). Your environment's specific availability requirements should dictate what replication factor you use.

Table 3: Recommendations for Replication Factor

| | |
|------------------------------------|---|
| Use replication factor 2 (default) | Suitable for all workload types. Suitable for all performance requirements. |
| Use replication factor 3 | Suitable for clusters with 32 or more nodes or as needed to meet availability requirements. |

5.1. DEAL WITH FAILURES

The DSF not only withstands a variety of hardware failures, it also builds strong redundancy into the software stack. Nutanix software processes that encounter a serious error are designed to fail fast. This design principle quickly restarts normal operations instead of waiting for a potentially faulty process to complete. Because the DSF continuously monitors components, it can stop and restart them when an error occurs to recover as quickly as possible, rather than letting them linger in an unresponsive state. Each host relies on its local CVM to service all storage requests. The DSF continuously monitors the health of all CVMs in the cluster. If an unrecoverable error occurs on a particular CVM, Nutanix autopathing automatically reroutes requests from the host to a healthy CVM on another node, providing data path redundancy.

This redirection continues until the local CVM failure issue is resolved. Because the cluster has a global namespace and access to replicas for all the data on that node, it can service requests immediately. This ability provides a high degree of fault tolerance and failover for all VMs in a Nutanix cluster. If the node's CVM continues to be unavailable for a prolonged period, data automatically replicates to maintain the necessary replication factor.

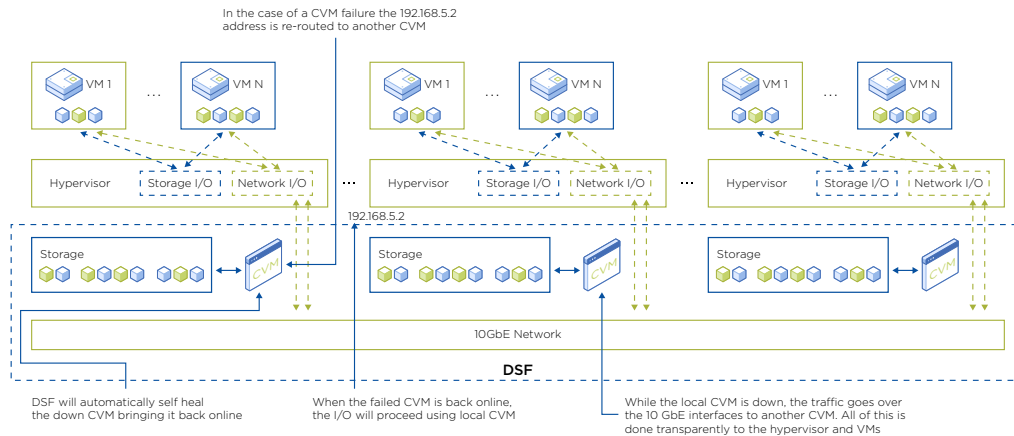


Figure 14: Data Path Redundancy

5.2. PREVENT NETWORK PARTITION ERRORS

The DSF uses the Paxos algorithm to avoid split-brain scenarios. Paxos is a proven protocol for reaching consensus or quorum among several participants in a distributed system. Before any file system metadata is written to Cassandra, Paxos ensures that all nodes in the system agree on the value. If the nodes do not reach a quorum, the operation fails in order to prevent any potential corruption or data inconsistency. This design protects against events like network partitioning, where communication between nodes may fail or packets may become corrupt, leading to a scenario where nodes disagree on values. The DSF also uses time stamps to ensure that updates are applied in the proper order.

5.3. PROACTIVELY RESOLVE BAD DISK RESOURCES

The DSF incorporates a Curator process that performs background housekeeping tasks to keep the entire cluster running smoothly. Among Curator's multiple responsibilities is ensuring file system metadata consistency and combing the extent store for corrupt and underreplicated data.

Additionally, Curator scans extents in successive passes, computes each extent's checksum, and compares it with the metadata checksum to validate consistency. If the checksums don't match, the corrupted extent is replaced with a valid extent from another node. This proactive data analysis protects against data loss and identifies bad sectors you can use to detect disks that are about to fail.

5.4. MAINTAIN AVAILABILITY: DISK FAILURE

The Nutanix unified component Stargate receives and processes data. All read and write requests for a node are sent to the Stargate process on that node. The Hades service simplifies the break-fix procedures for disks and automates several tasks that previously required manual user actions. Hades helps fix failing devices before they become unrecoverable.

Once Stargate sees delays in responses to I/O requests to a disk, it marks the disk offline. Hades then automatically removes the disk from the data path and runs `smartctl` checks against it. If the checks pass, Hades marks the disk online and returns it to service. If the checks fail or if Stargate marks a disk offline three times in one hour (regardless of the `smartctl` check results), Hades automatically starts the EC2 removal process. Removing the EC2 instance triggers an API call to the cluster portal, which notifies the Nutanix Clusters portal. The Nutanix Clusters portal allocates a new instance, adds it to the cluster, and marks the EC2 instance with the unresponsive disk for removal. The cluster software automatically replicates the data on the bad EC2 instance to other instances, then deletes the bad EC2 instance.

5.5. MAINTAIN AVAILABILITY: AVAILABILITY ZONE FAILURE

Availability Zones go offline for a variety of reasons—issues with power, cooling, or networking as well as scheduled system maintenance. We need to ensure that your NCA instance meets your availability needs. To avoid downtime in AWS, protect your workloads with Nutanix data protection or Leap. The destination for data protection or Leap could be another on-prem cluster or another NCA instance in a different Availability Zone.



6. Deployment Models

As customer environments can vary greatly, we provide several example deployment models. Regardless of the model you use, there are a few general outbound requirements for deploying a Nutanix cluster in AWS on top of the existing ones that on-prem clusters use for support services. The following table shows the endpoints the Nutanix cluster needs to communicate with for a successful deployment.

Note: Many of the destinations listed here use DNS failover and load balancing. For this reason the IP address returned when resolving a specific domain may change rapidly. We can't provide specific IP addresses in lieu of domain names.

Table 4: Cluster Outbound to the Cluster Portal

| Source | Destination | Protocol |
|-------------------|---|-----------------|
| Management subnet | gateway-external-api-prod.frame.nutanix.com | tcp/443 (HTTPS) |
| Management subnet | clusters-agents.s3-us-west-2.amazonaws.com | tcp/443 (HTTPS) |

Table 5: Cluster Outbound to EC2

| Source | Destination | Protocol |
|-------------------|--|-----------------|
| Management subnet | ec2.<region>.amazonaws.com (for example, a cluster in us-west-2 requires ec2.us-west-2.amazonaws.com) | tcp/443 (HTTPS) |
| Management subnet | aws.amazon.com | tcp/443 (HTTPS) |

Table 6: Cluster Outbound to Python Package Index

| Source | Destination | Protocol |
|-------------------|------------------------|-----------------|
| Management subnet | bootstrap.pypa.io | tcp/443 (HTTPS) |
| Management subnet | pip.pypa.io | tcp/443 (HTTPS) |
| Management subnet | pypi.org | tcp/443 (HTTPS) |
| Management subnet | files.pythonhosted.org | tcp/443 (HTTPS) |

General firewall support requirements are listed in the [Nutanix Security Guide](#).

6.1. MULTICLUSTER DEPLOYMENT

To protect your NCA cluster in the event of an Availability Zone failure, use your existing on-prem Nutanix Clusters instance as a disaster recovery target. There are many options when it comes to Nutanix disaster recovery, but here we're focusing on the native data protection included with every base Nutanix license.

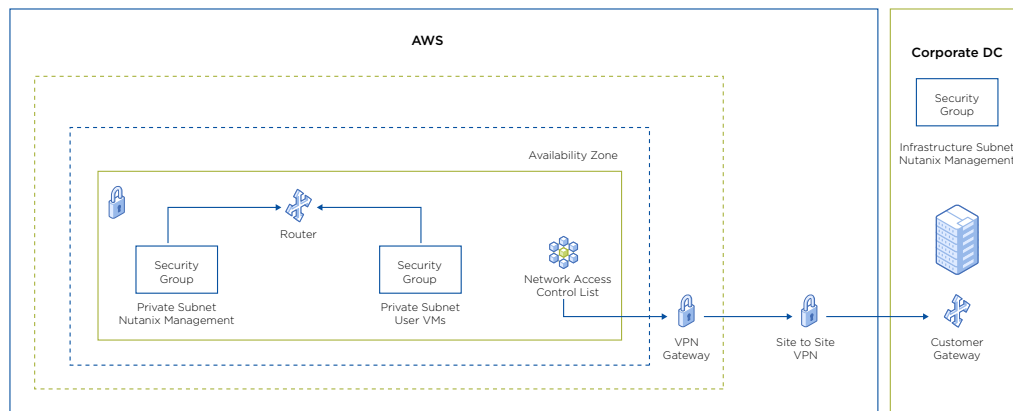


Figure 15: Multicloud Deployment

The following table details the inbound ports you need in order to establish replication between an on-prem cluster and a Nutanix cluster running in AWS. You can create these ports on the infrastructure subnet security group that was automatically created when you deployed NCA. The ports need to open in both directions.

Table 7: Inbound Security Group Rules for AWS

| Type | Protocol | Port Range | Source | Description |
|-----------------|----------|------------|--------------------|-----------------------|
| SSH | TCP | 22 | On-prem CVM subnet | SSH into the AHV node |
| Custom TCP rule | TCP | 2222 | On-prem CVM subnet | SSH access to the CVM |
| Custom TCP rule | TCP | 9440 | On-prem CVM subnet | UI access |
| Custom TCP rule | TCP | 2020 | On-prem CVM subnet | Replication |
| Custom TCP rule | TCP | 2009 | On-prem CVM subnet | Replication |

Make sure you set up the cluster virtual IP address for your on-prem and AWS clusters. This IP address is the destination address for the remote site.



The screenshot shows a 'Cluster Details' window with the following information:

| Field | Value |
|----------------------------|---|
| Cluster UUID | 00057368-cd89-acf2-0000-0000000088d1 |
| Cluster ID | 00057368-cd89-acf2-0000-0000000088d1::35025 |
| Cluster Incarnation ID | 1534268845698290 |
| Cluster Name | On-prem1 |
| Cluster Virtual IP Address | 10.19.170.29 |
| iSCSI Data Services IP | 10.19.170.30 |

Figure 16: Cluster Details

Near-synchronous (NearSync) and asynchronous replication are both valid options for customers running AHV on-prem. You can set your recovery point objective (RPO) to be as little as one minute with NearSync and one hour with asynchronous replication. If you want to use NearSync, your on-prem cluster needs to meet the requirements listed in the Prism Element Data Protection Guide.

6.2. MULTIPLE-AVAILABILITY ZONE DEPLOYMENT

If you don't have an on-prem cluster available for data protection or you want to use the low-latency links between Availability Zones, you can create a second NCA in a different Availability Zone. With this method, there is no data transfer charge between Amazon EC2 and other Amazon Web Services in the same AWS Region (for example, between Amazon EC2 US West and Amazon S3 US West), which can be a significant benefit.

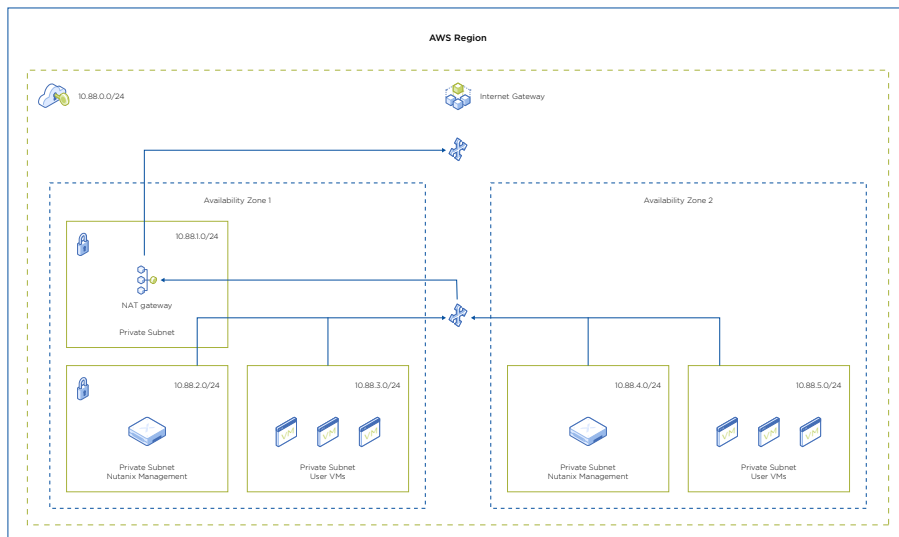


Figure 17: Multiple-Availability Zone Nutanix Clusters Deployment

Because you're still in your own custom VPC, your network design is very simple. You can isolate your private subnets for user VMs between clusters and use the private Nutanix management subnets to allow replication traffic between them. All private subnets can share the same routing table. Edit the inbound access in each Availability Zone's security group as shown in the following tables to allow replication traffic.

Table 8: Availability Zone 1 NCA Security Group settings

| Type | Protocol | Port Range | Source | Description |
|-----------------|----------|------------|--------------|-------------|
| Custom TCP Rule | TCP | 9440 | 10.88.4.0/24 | UI Access |
| Custom TCP Rule | TCP | 2020 | 10.88.4.0/24 | Replication |
| Custom TCP Rule | TCP | 2009 | 10.88.4.0/24 | Replication |

Table 9: Availability Zone 2 NCA Security Group Settings

| Type | Protocol | Port Range | Source | Description |
|-----------------|----------|------------|--------------|-------------|
| Custom TCP Rule | TCP | 9440 | 10.88.2.0/24 | UI Access |
| Custom TCP Rule | TCP | 2020 | 10.88.2.0/24 | Replication |
| Custom TCP Rule | TCP | 2009 | 10.88.2.0/24 | Replication |

Create route table

Actions

Filter by tags and attributes or search by keyword

| | Name | Route Table ID | Explicit subnet association | Edge associations | Main | VPC ID |
|--|--------------------------------------|------------------------------|-----------------------------|-------------------|------|------------------------------------|
| | | rtb-042e16dc4e3796b16 | subnet-0c59a3f136ef9c395 | - | Yes | vpc-0f8a2e8f634ac7e68 NutanixVPC |
| | XiCluster private route table | rtb-06f224be1444a1b59 | 5 subnets | - | No | vpc-0f8a2e8f634ac7e68 NutanixVPC |
| | | rtb-25ec184c | - | - | Yes | vpc-10c03579 |

Route Table: rtb-06f224be1444a1b59

Summary

Routes

Subnet Associations

Edge Associations

Route Propagation

Tags

Edit routes

View All routes

| Destination | Target | Status |
|--|-----------------------|--------|
| 10.88.0.0/16 | local | active |
| pl-6ea54007 (com.amazonaws.eu-central-1.s3, 52.219.44.0/22, 54.231.192.0/20, 52.219.72.0/22) | vpc-0196d19202b028e7d | active |
| 0.0.0.0/0 | nat-019b4c2909795b515 | active |

Figure 18: Private Route Table

If Availability Zone 1 goes down, you can activate protected VMs on the cluster in Availability Zone 2. Once Availability Zone 1 comes back online, you can redeploy a Nutanix cluster in Availability Zone 1 and reestablish data protection. New clusters require full replication.

6.3. SINGLE-AVAILABILITY ZONE DEPLOYMENT

Deploying a single cluster in AWS is great for more ephemeral workloads where you want to take advantage of performance improvements and use the same automation pipelines you use on-prem. If you decide not to use native Nutanix data protection to back up your workloads to your on-prem cluster or a second NCA, you must use another method; if your Availability Zone fails, you are not guaranteed the same EC2 instances when it resumes operation. You can use AHV-compatible backup products to target S3 as the backup destination, and, depending on the failure mode you want to recover from, you can also replicate that S3 bucket to a different Region. HYCU and Veeam are two backup methods proven to work with AHV.

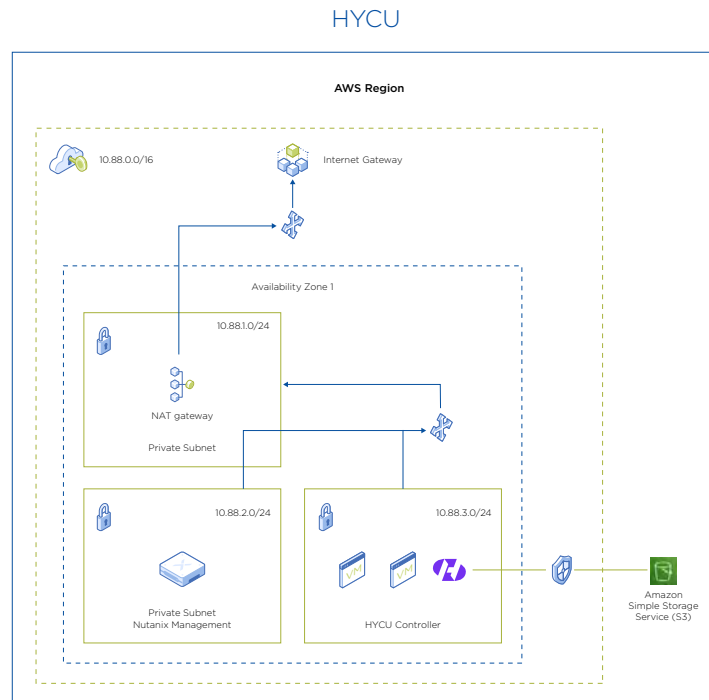


Figure 19: HYCU Backup Architecture

HYCU has the following limitations:

- HYCU does not support AWS S3 targets that use the Glacier storage class.
- HYCU currently only supports AWS S3 Signature version 4.

Deploy HYCU

The **HYCU backup controller** runs on the deployed Nutanix cluster as a VM. It needs to communicate with the private subnet to register with Prism Element. Once the HYCU controller is running on the Nutanix cluster, you can configure the S3 endpoint.

TIP: Deploy a VPC endpoint for S3 so your backup traffic doesn't go over the Internet.

The following procedure walks you through creating an S3 endpoint:

- Open the Amazon VPC console. In the navigation pane, choose **Endpoints**.
- The opened page asks you to create your first endpoint. Click **Create Endpoint**.
- Choose your VPC and specify a policy that controls access to the AWS service. Allow full access; if you don't, only IAM users can access the service.
 - If you want to create a locked user profile, you must specify at least these AWS S3 permissions: s3:GetObject, s3:DeleteObject, s3:PutObject, s3:ListBucket, s3:GetBucketAcl, s3:ListBucketMultipartUploads, and s3:GetBucketLocation.
- Associate the endpoint with your private subnet.

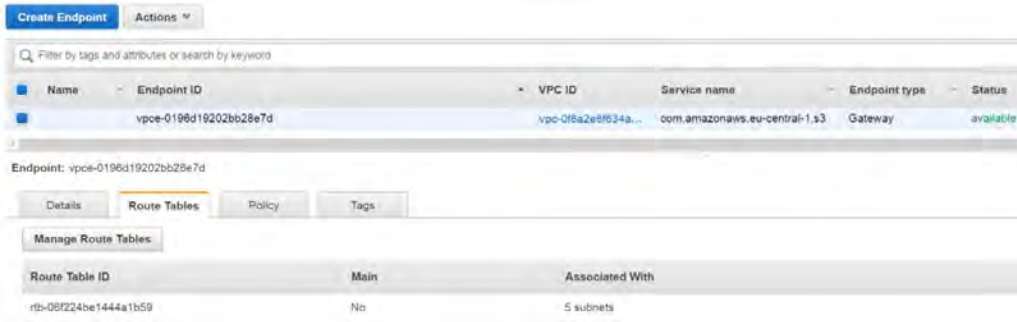


Figure 21: S3 VPC Endpoint

You can use the VPC endpoint to block all public access to your public S3 bucket.

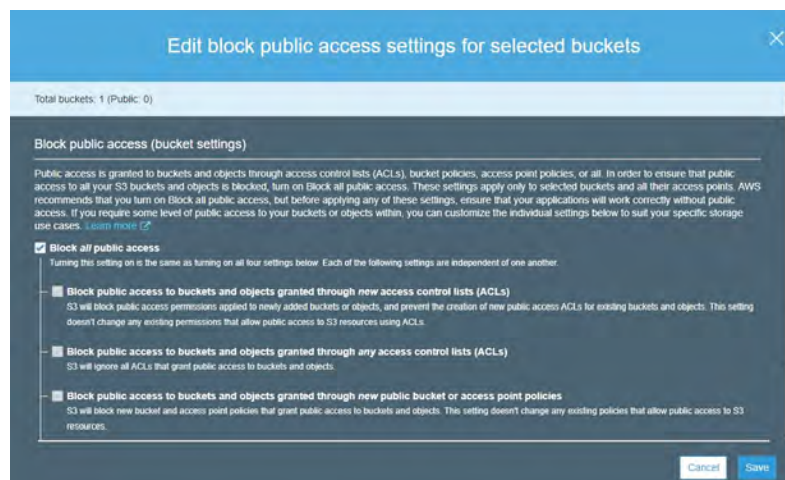


Figure 21: Block Public Access to Backup S3 Bucket

Recover from NCA Failure with HYCU

To ensure that you can restore your VMs to a new NCA instance in the case of unrecoverable NCA failure:

- Back up the VMs and the HYCU controller to a S3 target.
- Once you deploy a new NCA instance, spin up a temporary HYCU controller on that cluster.
- Before you import existing targets, suspend all other activities on the controller.
- For faster restores, put the HYCU controller backup in its own S3 bucket.

Follow these steps to restore your VMs to the new NCA:

- Deploy a temporary HYCU backup controller.
 - Import the targets that store the backup of the original HYCU backup controller.
 - Add the new NCA as the target location for restoring your HYCU backup controller.
- If you plan to restore VMs and applications, add locations for those as well.

You can find more detailed instructions for restoring VMs in the [HYCU support documentation](#).

Recover from Region Failure with HYCU

Replicate your S3 bucket to a different region. AWS provides cross-Region replication (CRR) to copy objects in Amazon S3 buckets across Regions. With this tool, you can deploy the NCA in the new Region and replicate your S3 bucket to the original site once the source Region comes back. To use CRR, you need to enable versioning on your source and destination S3 buckets, as shown in the following image.

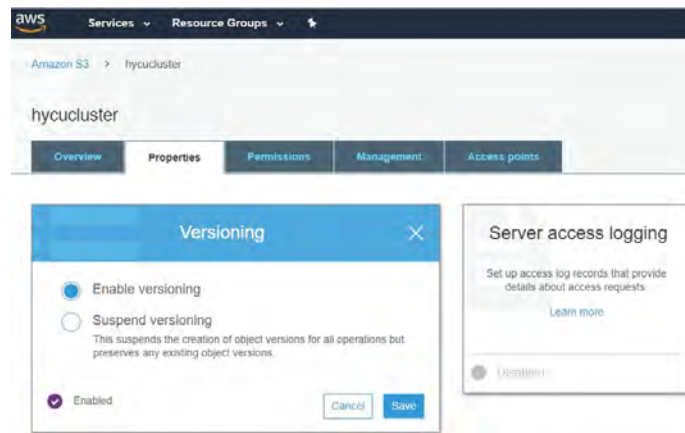


Figure 22: Enable Versioning to Use CRR

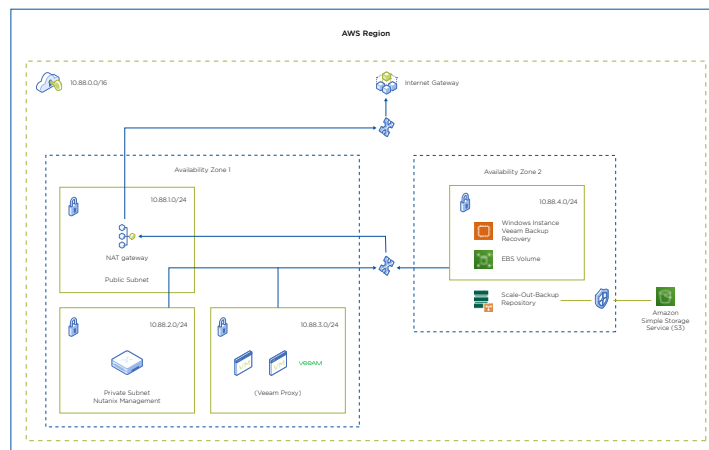


Figure 23: Veeam Backup Architecture

Deploy Veeam Backup & Replication (VBR)

You can deploy Veeam for NCA several ways but regardless of the overall deployment strategy, you must run a Windows EC2 instance because Veeam requires all block storage in the scale-out backup repository to use S3. For this reason you need to have block storage and S3 available to create the scale-out repository. You must also deploy an AHV proxy connected to VBR and Prism Element on NCA. You could also run Veeam Backup & Replication (VBR) directly as a VM on NCA and use a Linux EC2 appliance as the backup repository.

Table 10: Veeam Ports between the Nutanix Cluster and VBR

| From | To | Port | Protocol | Description |
|--|---|-----------|------------|--|
| Browser | Veeam Availability for Nutanix AHV server | 8100 | HTTPS | Web UI of the proxy appliance. |
| Veeam Availability for Nutanix AHV (proxy appliance) | Nutanix REST API | 9440 | HTTPS | Port used to connect with Nutanix REST API. |
| | Veeam Backup & Replication server | 10006 | TCP | Port used to connect to Veeam Backup & Replication. |
| | Nutanix AHV server | 3260 | TCP/ iSCSI | For connecting to disks on Nutanix AHV. |
| | Veeam Agent server | 2500-5000 | TCP | Default range of ports used as transmission channels for jobs and restore sessions. For every TCP connection a job uses, one port from this range is assigned. |

To see all the ports used for Veeam Backup & Replication, backup proxy, and backup repositories, see the Used Ports section of the Veeam Backup & Replication User Guide.

The AWS S3 setup for your bucket is the same as the configuration when using HYCU, with the following exceptions:

- AWS S3 only supports the Standard, Standard-Infrequent Access, and One Zone-IA storage classes. Use S3 Standard to store frequently accessed data, and use S3 Standard-IA and S3 One Zone-IA to store long-lived but less frequently accessed data. Choose your storage class based on your requirements.
- Don't use any mechanism other than VBR to manage data and data retention in an object storage bucket or container.
- VBR doesn't support enabling life cycle rules; doing so may result in backup and restore failures.

Use the VPC endpoint to keep your S3 public bucket blocked from any public access.

Recover from Availability Zone Failure with VBR

If you run VBR in a different Availability Zone than your NCA, it's easy to recover if the Availability Zone where NCA runs fails. To use VBR to recover from an Availability Zone failure, deploy a new Veeam AHV proxy and register it with VBR. We recommend keeping a configuration backup of the Veeam AHV proxy. If you decide to keep a configuration backup of the proxy, use the settings in the following figure.

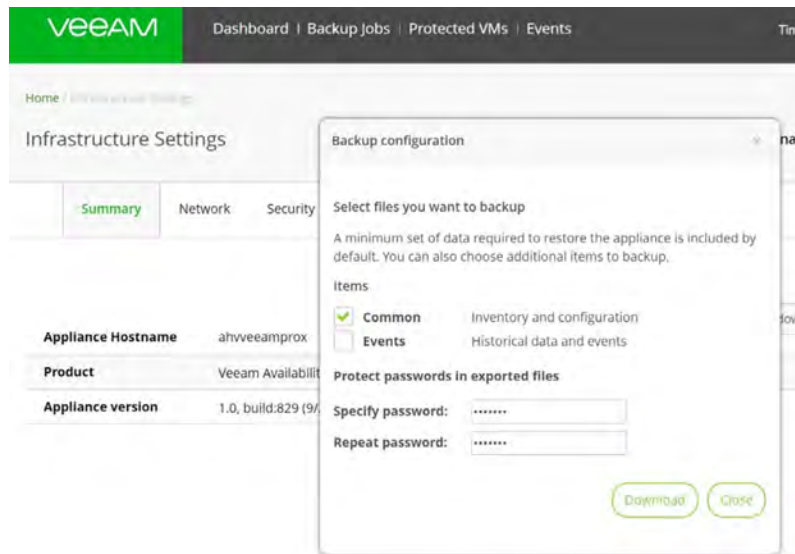


Figure 24: Backup Configuration of AHV Proxy

The following high-level instructions walk you through restoring your NCA instance after an Availability Zone failure:

- Deploy a new Veeam proxy to the new NCA cluster.
- Register the new Veeam proxy with the NCA cluster.
- In the Veeam Backup & Replication console, scan the repository.
- In the Veeam proxy appliance console, click the gear icon, click Manage Veeam Servers, then select the required Veeam backup server from the list.
- Click Import Backups, then click Proceed to confirm the action. The proxy appliance scans the backup repositories and imports all compatible backups of AHV VMs.
- Restore the imported backups.


Recover from Region Failure with VBR

To recover from a Region failure, you must have an additional scale-out backup repository set up in a different AWS Region using a Linux-based EC2 VM. Because the scale-out backup uses Elastic Block Storage (EBS) and S3, retention rules in the backup policy don't guarantee that all data has been moved to S3. Run a backup copy job after your backup to ensure that you get the latest copies:

- In the VBR console, click **New Backup Copy Job**.
- In the window that opens, click **Target**. Select the correct backup repository from the dropdown menu and enter 7 in Restore points to keep. Click **Next**.
- Define your backup window and finish the configuration.

New Backup Copy Job

×



Target
Specify the target backup repository, amount of most recent restore points to keep, and retention policy for full backups. You can use map backup functionality to seed the backup files.

Job

Objects

Target

Data Transfer

Schedule

Summary

Backup repository:

VeeamRA (Created by EC2AMAZ-OGP8OKA\Administrator at 3/4/2020 3:54 PM.)

217 GB free of 599 GB

[Map backup](#)

Restore points to keep: 7

☐ Keep the following restore points as full backups for archival purposes

Weekly backup: 4

Saturday

Schedule

Monthly backup: 0

First Sunday of the month

Quarterly backup: 0


First Sunday of the quarter

Yearly backup: 0

First Sunday of the year

☐ Read the entire restore point from source backup instead of synthesizing it from increments

Advanced settings include health check and compact schedule, notifications settings, and automated post-job activity options.


Advanced

< Previous

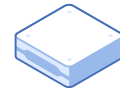
Next >

Finish

Cancel

Figure 25: Backup Copy Job

©2020 NUTANIX, INC. ALL RIGHTS RESERVED | 35



7. Capacity Optimization

The Nutanix hybrid multicloud software offers capacity optimization features that improve storage utilization and performance. The two key features are compression and deduplication.

7.1. COMPRESSION

Nutanix systems currently offer two types of compression policies, as described in the following table.

Table 11: Nutanix Compression Policies

| | |
|--------------|--|
| Inline | The system compresses data synchronously as it is written to optimize capacity and to maintain high performance for sequential I/O operations. Inline compression only compresses sequential I/O to avoid degrading performance for random write I/O. |
| Post-process | For random workloads, data writes to the SSD tier uncompressed for high performance. Compression occurs after cold data migrates to lower-performance storage tiers. Post-process compression acts only when data and compute resources are available, so it doesn't affect normal I/O operations. |

Nutanix recommends that all customers carefully consider the advantages and disadvantages of compression for their specific applications. For further information on compression, please refer to the [Nutanix Data Efficiency tech note](#).

7.2. DEDUPLICATION

The software-driven Elastic Deduplication Engine increases the effective capacity in the disk tier, as well as the utilization of the performance tiers (RAM and flash), by eliminating duplicate data. By providing for larger effective cache sizes in the performance tier, this feature substantially increases performance for certain workloads.

Deduplication savings vary greatly depending on workload and data types, but, in general, deduplication provides the largest benefit for common data sets, such as full-clone VDI workloads. Nutanix does not recommend deduplication for general-purpose server workloads, including business-critical applications.

The following table offers general guidelines but Nutanix recommends that all customers carefully consider the advantages and disadvantages of deduplication for their specific applications. For further information on deduplication, please refer to the Nutanix Data Efficiency tech note

Table 12: Recommendations for Deduplication per Container (Datastore)

| | |
|---|--|
| Containers hosting business-critical applications | Disable deduplication for all except full-clone VDI VMs. Be sure to increase CVM memory to at least 24 GB. |
| Containers hosting VDI | |
| Containers hosting general server workloads | |
| Containers hosting big data | |



8. Encryption

To help reduce cost and complexity, Nutanix supports a native local key manager (LKM) for all clusters with three or more nodes. The LKM runs as a service distributed among all the nodes. You can activate it easily from Prism Element to enable encryption without adding another silo to manage. Customers looking to simplify their infrastructure operations can now have one-click infrastructure for their key manager as well.

Organizations often purchase external key managers (EKMs) separately for both software and hardware. However, because the Nutanix LKM runs natively in the CVM, it's highly available and there is no variable add-on pricing based on the number of nodes. Every time you add a node you know the final cost. You also gain peace of mind because when you upgrade your cluster, the key management services are also upgraded. When upgrading the infrastructure and management services in lockstep, you're ensuring your security posture and availability by staying in line with the support matrix.

Nutanix software encryption provides native AES-256 data-at-rest encryption, which can interact with any KMIP- or TCG-compliant external KMS server (Vormetric, SafeNet, etc.) and the native Nutanix native KMS, introduced in AOS version 5.8. The system uses Intel AES-NI acceleration for encryption and decryption processes to minimize any potential performance impacts.

We recommend using the native Nutanix KMS to provide additional security for your workloads in the cloud.

NOTE: The first copy of the data (written locally) is encrypted. The copy sent over the wire is also encrypted and stored on a remote node.



9. Virtual Machine High Availability

VM high availability (VMHA) ensures that VMs restart on another AHV host in the cluster if a host fails. VMHA considers RAM when calculating available resources throughout the cluster for starting VMs.

VMHA respects affinity and antiaffinity rules. For example, with VM-host affinity rules, VMHA does not start a VM pinned to AHV host 1 and host 2 on another host when those two are down unless the affinity rule specifies an alternate host.

There are two VM high availability modes:

- **Default**
This mode requires no configuration and is included by default when you deploy an AHV- based Nutanix cluster. When an AHV host becomes unavailable, the VMs that were running on the failed AHV host restart on the remaining hosts, depending on the available resources. If the remaining hosts do not have sufficient resources, some of the failed VMs may not restart.
- **Guarantee**
This nondefault configuration reserves space throughout the AHV hosts in the cluster to guarantee that all VMs can restart on other hosts in the AHV cluster during a host failure. To enable Guarantee mode, select the Enable HA check box, as shown in the following figure. A message then displays the amount of memory reserved and how many AHV host failures the system can tolerate.

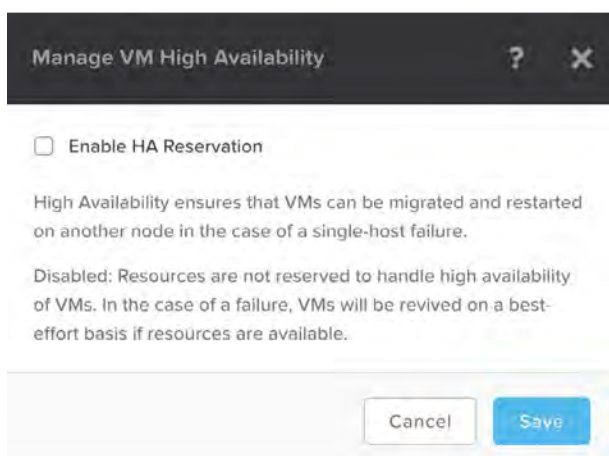


Figure 26: VM High Availability Reservation

The VMHA configuration reserves resources to protect against:

- One AHV host failure, if all Nutanix containers are configured with replication factor 2.
- Two AHV host failures, if any Nutanix container is configured with replication factor 3.

Admins can use the aCLI to manage protection against two AHV host failures when using replication factor 3. Use the following command to designate the maximum number of tolerable AHV host failures:

```
nutanix@CVM$ acli ha.update num_host_failures_to_tolerate=X
```

When an unavailable AHV host comes back online after a VMHA event, VMs previously running on that host migrate back to maintain data locality.

To disable VMHA per VM, set a negative value (-1) when creating or updating the VM. This configuration removes the VM from the VMHA resource calculation.

```
nutanix@CVM$ acli vm.update <VM Name> ha_priority=-1
nutanix@CVM$ acli vm.create <VM Name> ha_priority=-1
```

In this configuration, the VM does not start on a new AHV host when its host fails; it only starts again when the failed host comes back online.

9.1. VMHA RECOMMENDATIONS

- Use the nondefault VMHA Guarantee mode when you need to ensure that all VMs can restart if an AHV host fails.
- When using Guarantee mode, keep the default reservation type of `kAcropolisHAReserveSegments`; this setting must not be altered.

NOTE: The VMHA reservation type `kAcropolisHAReserveHosts` is deprecated. Never change the VMHA reservation type to `kAcropolisHAReserveHosts`.

- Consider storage availability requirements when using VMHA Guarantee mode. Ensure that the parameter `num_host_failures_to_tolerate` is less than the configured storage availability. If there are only two copies of the VM data, the VM data could be unavailable if two hosts are down at the same time even though there are CPU and RAM resources to run the VMs.
- You have to disable VMHA before you can use the Acropolis Dynamic Scheduler (ADS) VM- host affinity feature to pin a VM to one AHV host. However, we don't recommend pinning VMs to a particular AHV host, as we discuss in the following section.



10. Acropolis Dynamic Scheduler

ADS ensures that compute (CPU and RAM) and storage resources are available for VMs and volume groups (VGs) in the Nutanix cluster. You also use ADS to define affinity policies.

10.1. AFFINITY POLICIES

You define affinity policies one of two ways: manually (if you're a Nutanix administrator) or with a VM-provisioning workflow. There are two affinity policies:

- VM-host affinity

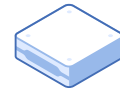
This configuration keeps a VM on a specific set of AHV hosts. It is useful when you need to limit VMs to a subset of available AHV hosts because of application licensing, host resources (such as available CPU cores or CPU GHz speed), available RAM or RAM speed, or local SSD capacity. Host affinity is a must rule: AHV always honors the specified rule.

NOTE: We recommend against using VM-host affinity.

- VM-VM antiaffinity

This configuration ensures that two or more VMs do not run on the same AHV host. It is useful when an application provides high availability and an AHV host must not be the application's single point of failure. Antiaffinity is a should rule that is honored only when there are enough resources available to run VMs on separate hosts.

For additional information about ADS and affinity policies, read the [ADS section](#) of the [AHV best practices guide](#).



11. Conclusion

Nutanix software running in the cloud provides an easy extension for your on-premises datacenter. If you're already consuming cloud resources, the native Nutanix integration with AWS means that you don't need any additional skills to get your workloads running in the cloud. Management overhead shrinks when you no longer need an additional overlay network to secure and lock down networking between your on-prem environment and the cloud. Once you have Nutanix Clusters running in the cloud, you can enjoy native networking speeds between migrated workloads and the new cloud services you want to consume. For more information, check out the [Nutanix Clusters website](#).



info@nutanix.com | www.nutanix.com |  @nutanix

Nutanix makes infrastructure invisible, elevating IT to focus on the applications and services that power their business. The Nutanix Enterprise Cloud OS leverages web-scale engineering and consumer-grade design to natively converge compute, virtualization, and storage into a resilient, software-defined solution with rich machine intelligence. The result is predictable performance, cloud-like infrastructure consumption, robust security, and seamless application mobility for a broad range of enterprise applications. Learn more at www.nutanix.com or follow us on [Twitter @nutanix](https://twitter.com/nutanix).