Nutanix Cloud Clusters - Nutanix Cloud Clusters on AWS

[PDF generated June 03 2025. For all recent updates please see the Nutanix Bible releases notes located at https://nutanixbible.com/release_notes.html. Disclaimer: Downloaded PDFs may not always contain the latest information.]

Nutanix Cloud Clusters (NC2) on AWS provides on-demand clusters running in target cloud environments using bare metal resources. This allows for true on-demand capacity with the simplicity of the Nutanix platform you know. Once provisioned the cluster appears like any traditional AHV cluster, just running in a cloud providers datacenters.

Supported Configurations

The solution is applicable to the configurations below (list may be incomplete, refer to documentation for a fully supported list):

Core Use Case(s):

- On-Demand / burst capacity
- Backup / DR
- Cloud Native
- Geo Expansion / DC consolidation
- App migration

Management interfaces(s):

- Nutanix Clusters Portal - Provisioning
- Prism Central (PC) - Nutanix Management
- AWS Console - AWS Management

- EC2 Metal Instance Types:
  - i3.metal
  - m5d.metal
  - z1d.metal
  - i3en.metal
  - g4dn.metal
  - i4i.metal
  - m6d.metal

Upgrades:

- Part of AOS

Compatible Features:

- AOS Features
- AWS Services

Key terms / Constructs

The following key items are used throughout this section and defined in the following:

- Nutanix Clusters Portal
  - The Nutanix Clusters Portal is responsible for handling cluster provisioning requests and interacting with AWS and the provisioned hosts. It creates cluster specific details and handles the dynamic CloudFormation stack creation.

- Region
  - A geographic landmass or area where multiple Availability Zones (sites) are located. A region can have two or more AZs. These can include regions like US-East-1 or US-West-1.

- Availability Zone (AZ)
  - An AZ consists of one or more discrete datacenters inter-connected by low latency links. Each site has it's own redundant power, cooling, network, etc. Comparing these to a traditional colo or datacenter, these would be considered more resilient as a AZ can consist of multiple independent datacenters. These can include sites like US-East-1a or US-West-1a.

- VPC
  - A logically isolated segment of the AWS cloud for tenants. Provides a mechanism to to secure and isolate environment from others. Can be exposed to the internet or other private network segments (other VPCs, or VPNs).

- S3
  - Amazon's object service which provides persistent object storage accessed via the S3 API. This is used for archival / restore.

- EBS
  - Amazon's volume / block service which provides persistent volumes that can be attached to AMIs.

- Cloud Formation Template (CFT)
  - A Cloud Formation Template simplifies provisioning, but allowing you to define a "stack" of resources and dependencies. This stack can then be provisioned as a whole instead of each individual resource.

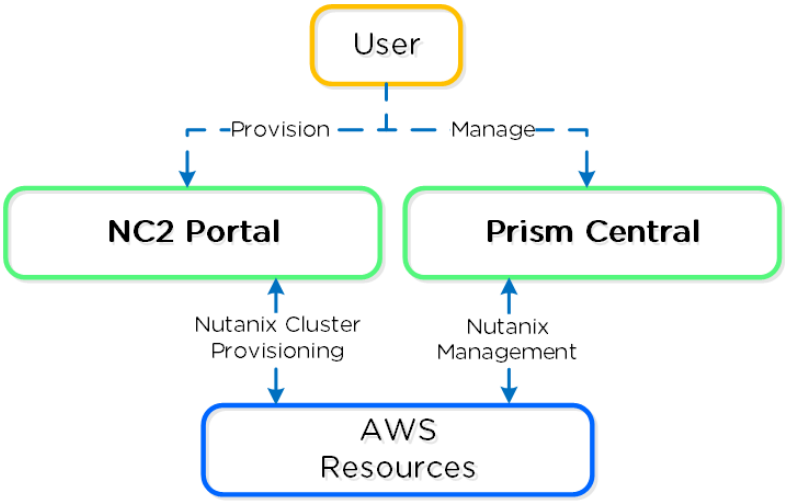
Cluster Architecture

From a high-level the Nutanix Clusters Portal is the main interface for provisioning Nutanix Clusters on AWS and interacting with AWS.

The provisioning process can be summarized with the following high-level steps:

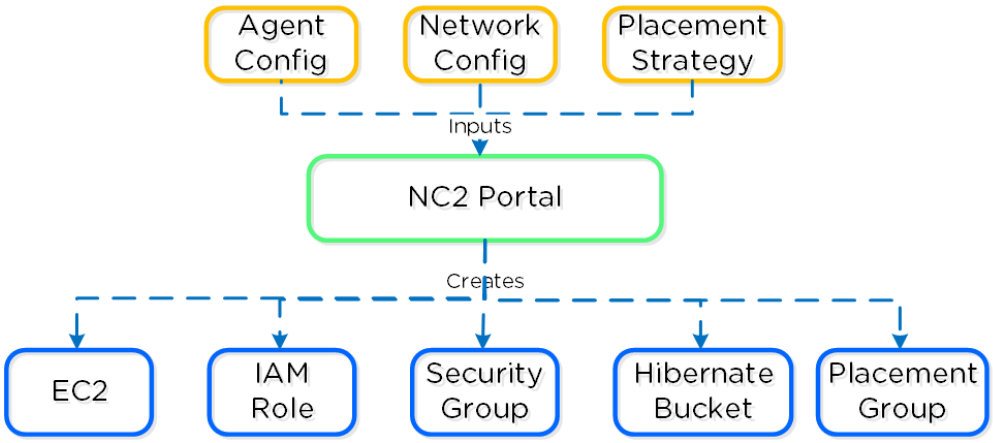
1. Create cluster in NC2 Portal
2. Deployment specific inputs (e.g. Region, AZ, Instance type, VPC/Subnets, etc.)
3. The NC2 Portal creates associated resources
4. Host agent in Nutanix AMI checks-in with Nutanix Clusters on AWS
5. Once all hosts as up, cluster is created

The following shows a high-level overview of the NC2A interaction:



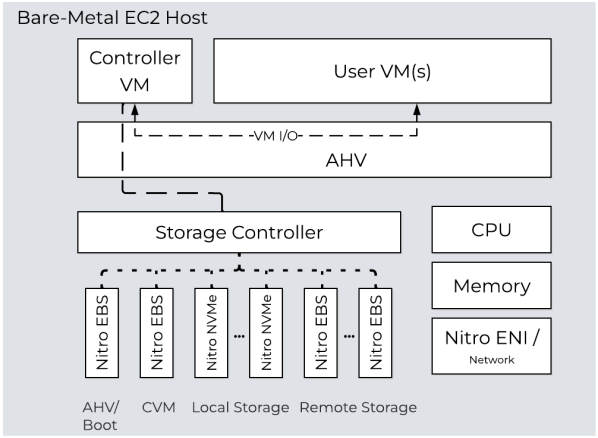
NC2A - Overview

The following shows a high-level overview of a the inputs taken by the NC2 Portal and some created resources:



Nutanix Clusters on AWS - Cluster Orchestrator Inputs

The following shows a high-level overview of a node in AWS:



NC2A - Node Architecture

Storage for Nutanix Cloud Clusters on Amazon Web Services

The primary storage for NC2 comes from the locally attached NVMe disks. The locally attached disks are the first tier that the CVMs use to persist the data. Each AWS node also consumes two AWS Nitro EBS volumes that are attached to the bare-metal node. One of those EBS volumes is used for AHV and the other for the CVM. When you provision the NC2 instance in the NC2 portal, you can add more Nitro EBS volumes to each bare-metal node in the cluster as remote storage, scaling your storage to meet business needs without adding more bare-metal nodes. If you initially create the cluster with Nitro EBS volumes, you can add more Nitro EBS volumes later.

Given the hosts are bare metal, we have full control over storage and network resources similar to a typical on-premises deployment. For the CVM and AHV host boot, EBS volumes are used. NOTE: certain resources like EBS interaction run through the AWS Nitro card which appears as a NVMe controller in the AHV host.

When you add EBS volumes to a cluster, they're added in a uniform manner throughout the cluster. The total amount of storage follows the same AOS limitations (Nutanix portal credentials required) in terms of the maximum amount storage that can be added, with the additional constraint that the maximum storage can't be more than 20 percent of the local storage for the bare-metal node. If you use the I4i.metal instance, which has 30 TB of local AWS Nitro NVMe-based storage, you can reach the current AOS node limit with the other 80 percent when using snapshots.

Through the NC2 portal, you can add a small amount of EBS storage during the initial deployment and add more later. If you have a disaster recovery use case with NC2, you can deploy a three-node cluster with a small amount of EBS storage to cover the storage usage for tier-1 workloads that need fast recovery. If your needs change, you can scale up the storage later. At failover, you

can use Prism Central playbooks or the NC2 portal to add NC2 nodes to cover any RAM usage not supplied by the three bare-metal nodes.

Nutanix recommends keeping the minimum number of additional nodes greater than or equal to your cluster's redundancy factor. Expand the cluster in multiples of your redundancy factor for the additional nodes.

Note: If you use EBS volumes for storage, only use homogenous clusters (all the nodes must be the same type).

The data tiering process for NC2 on AWS is the same as the process for a hybrid configuration on-premises.

Placement policy

Nutanix uses a partition placement strategy when deploying nodes inside an AWS Availability Zone. 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.

NC2 on AWS supports combining heterogenous node types in a cluster. You can deploy a cluster of one node type and then expand that cluster’s capacity by adding heterogenous nodes to it. This feature protects your cluster if its original node type runs out in the Region and provides flexibility when expanding your cluster on demand. If you're looking to right-size your storage solution, support for heterogenous nodes can give you more instance options to choose from.

NC2 on AWS allows you to use a combination of i3.metal, i3en.metal, and i4i.metal instance types or z1d.metal, m5d.metal, and m6id.metal instance types while creating a new cluster or expanding the cluster capacity of an already running cluster.

NC2 on AWS allows you to create a heterogeneous cluster depending on the following conditions:

NC2 on AWS supports a combination of i3.metal, i3en.metal, and i4i.metal instance types or z1d.metal, m5d.metal, and m6id.metal instance types. The AWS region must have these instance types supported by NC2 on AWS. For more information, see Supported Regions and Bare-metal Instances.

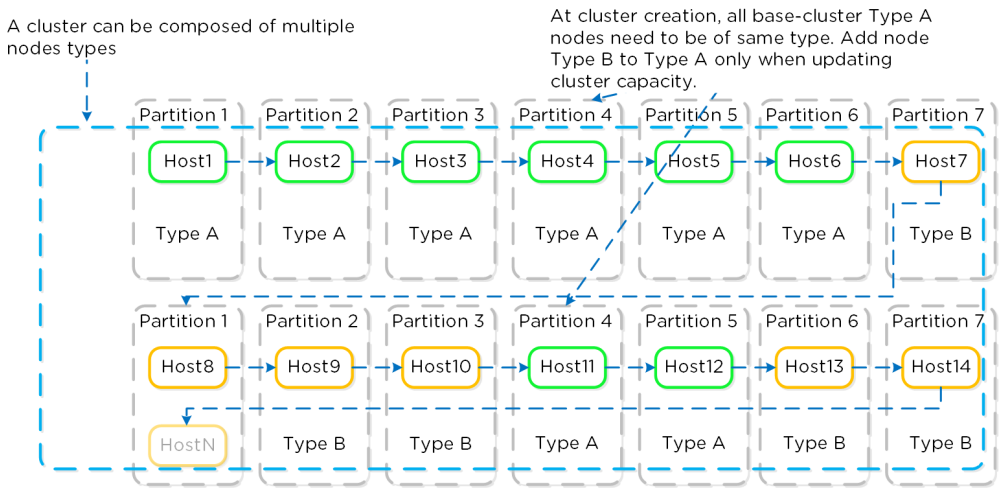
Note: You can only create homogenous clusters with g4dn.metal; it cannot be used to create a heterogeneous cluster.

Nutanix recommends that the minimum number of additional nodes must be equal to or greater than your cluster's redundancy factors (RF), and the cluster must be expanded in multiples of RF for the additional nodes. A warning is displayed if the number of nodes is not evenly divisible by the RF number.

UVMs that have been created and powered ON in the original cluster running a specific node or a combination of compatible nodes, as listed below, cannot be live migrated across different node types when other nodes are added to the cluster. After successful cluster expansion, all UVMs must be powered OFF and powered ON to enable live migration.

If z1d.metal is present in the heterogeneous cluster either as the initial node type of the cluster or as the new node type added to an existing cluster. If i4i.metal is the initial node type of the cluster and any other compatible node is added. If m6id.metal is the initial node type of the cluster and any other compatible node is added. If i3en.metal is the initial node type of the cluster and the i3.metal node is added.

You can expand or shrink the cluster with any number of i3.metal, i3en.metal, and i4i.metal instance types or z1d.metal, m5d.metal, and m6id.metal instance types as long as the cluster size remains within the cap of a maximum of 28 nodes



NC2A - Partition Placement

When you've formed the Nutanix cluster, the partition groups map to the Nutanix rack-awareness feature. AOS Storage 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.

Storage

Storage for Nutanix Cloud Clusters on AWS can be broken down into two core areas:

1. Core / Active
2. Hibernation

Core storage is the exact same as you'd expect on any Nutanix cluster, passing the "local" storage devices to the CVM to be leveraged by Stargate.

Instance Storage

Given that the "local" storage is backed by the AWS instance store, which isn't fully resilient in the event of a power outage / node failure additional considerations must be handled.

For example, in a local Nutanix cluster in the event of a power outage or node failure, the storage is persisted on the local devices and will come back when the node / power comes back online. In the case of the AWS instance store, this is not the case.

In most cases it is highly unlikely that a full AZ will lose power / go down, however for sensitive workloads it is recommended to:

- Leverage a backup solution to persist to S3 or any durable storage
- Replicate data to another Nutanix cluster in a different AZ/Region/Cloud (on-prem or remote)

One unique ability with NC2A is the ability to "hibernate" a cluster allowing you to persist the data while spinning down the EC2 compute instances. This could be useful for cases where you don't need the compute resources and don't want to continue paying for them, but want to persist the data and have the ability to restore at a later point.

When a cluster is hibernated, the data will be backed up from the cluster to S3. Once the data is backed up the EC2 instances will be terminated. Upon a resume / restore, new EC2 instances will be provisioned and data will be loaded into the cluster from S3.

Networking

Networking can be broken down into a few core areas:

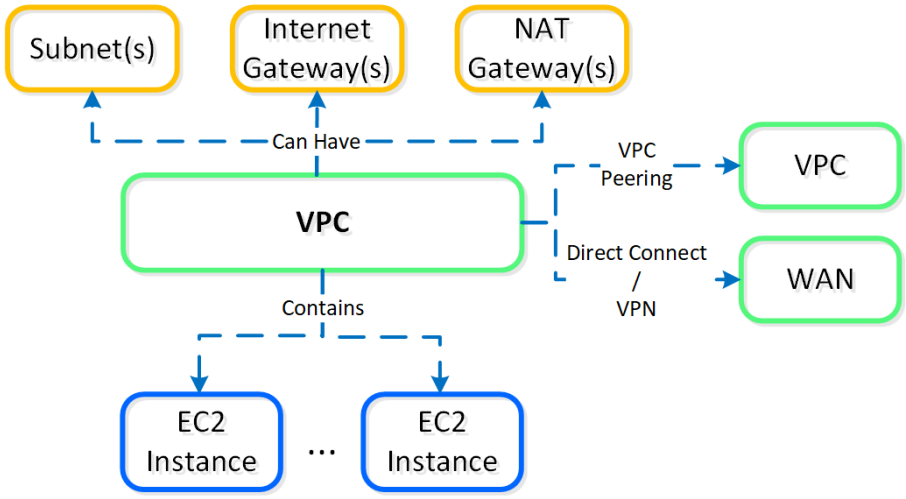
- Host / Cluster Networking
- Guest / UVM Networking

· WAN / L3 Networking

Native vs. Overlay

Instead of running our own overlay network, we decided to run natively on AWS subnets, this allows VMs running on the platform to natively communicate with AWS services with zero performance degradation.

NC2A are provisioned into an AWS VPC, the following shows a high-level overview of an AWS VPC:



NC2A - AWS VPC

New vs. Default VPC

AWS will create a default VPC/Subnet/Etc. with a 172.31.0.0/16 ip scheme for each region.

It is recommended to create a new VPC with associated subnets, NAT/Internet Gateways, etc. that fits into your corporate IP scheme. This is important if you ever plan to extend networks between VPCs (VPC peering), or to your existing WAN. This should be treated as you would treat any site on the WAN.

Nutanix Cloud Networking

Nutanix delivers a truly hybrid multicloud experience because you can use the standalone native cloud networking or Flow Virtual Networking, which gives you the option for the low overhead of native networking with the same operational model as the rest of your hybrid multicloud.

Native Networking

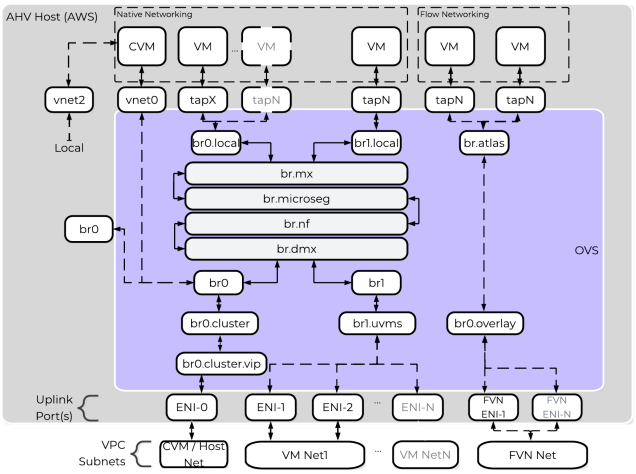
Nutanix integration with the AWS networking stack means that every VM deployed on NC2 on AWS receives a native AWS IP address when using native networking, so that applications have full access to all AWS resources as soon as you migrate or create them on NC2 on AWS. Because the Nutanix network capabilities are directly on top of the AWS overlay, network performance remains high and resource consumption is low because you don't need additional network components.

With native network integration, you can deploy NC2 in existing AWS Virtual Private Clouds (VPCs). Because existing AWS environments have gone through change control and security processes already, you only need to allow NC2 on AWS to talk to an NC2 portal. With this integration, you can increase security in your cloud environments.

Nutanix uses native AWS API calls to deploy AOS on bare-metal Amazon Elastic Compute Cloud (EC2) instances and consume network resources. Each bare-metal EC2 instance has full access to its bandwidth through an ENI, so if you deploy Nutanix to an i3.metal instance, each node has access to 25 Gbps. The ENI is a logical networking component in a VPC that represents a virtual network card. An ENI can have one primary IP address and up to 49 secondary IP addresses. AWS hosts can support up to 15

ENIs. An ENI is consumed each time you turn on a VM for a new subnet. If one particular host in one subnet has more than 49 VMs, the network service running on the NC2 host automatically adds an additional ENI to support the workload. While an NC2 instance has access to a configured AWS subnet on every host, VMs that aren't on don't consume ENIs. All deployed guest VMs use the secondary IP addresses to obtain direct AWS network access. AHV hosts deployed in AWS have separate ENIs for management traffic (AHV and CVM) and guest VMs, which means that you can have different AWS security groups for management and guest VMs. 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. Native networking uses the same networking stack as on-premises. Flow Virtual Networking uses separate ENIs to allow traffic to exit the cluster. The following figure shows a conceptual diagram of the OVS architecture.



NC2 on AWS - OVS Architecture

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

NC2 on AWS creates a single default security group for guest VMs running in the Nutanix cluster. Any ENIs created to support guest VMs are members of this default security group, which allows all guest VMs in a cluster to communicate with each other. In addition to security groups, you can use Nutanix Flow Network Security to provide greater security controls for east-west network traffic.

For UVM (Guest VM) networking, VPC subnets are used. A UVM network can be created during the cluster creation process or via the following steps:

From the AWS VPC dashboard, click on 'subnets' then click on 'Create Subnet' and input the network details:

Creating a Subnet

An AWS Region is a distinct geographic area. Each Region has multiple, isolated locations known as Availability Zones (AZs), which are logical datacenters available for any AWS customer in that Region to use. Each AZ in a Region has redundant and separate power, networking, and connectivity to reduce the likelihood of two AZs failing simultaneously.

Create a subnet in AWS in a VPC, then connect it to AOS in Prism Element. Cloud network, a new service in the CVM, works 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 prevents you from using the AHV or CVM subnet for guest VMs by not allowing you 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 that 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 that it needs for a successful deployment, you don't need to do any additional work to use the AWS overlay network.

Always keep the following best practices in mind:

- Don't share AWS guest-VM subnets between clusters.
- Have separate subnets for management (AHV and CVM) and guest 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 AZs in a Region.
- In each AZ, 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.

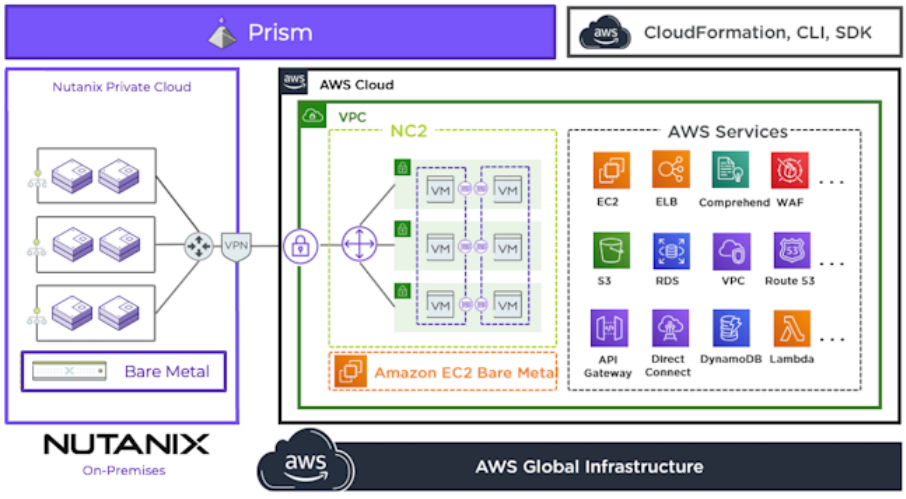
Guest AHV IP Address Management

AHV uses IP address management (IPAM) to integrate with native AWS networking. NC2 on AWS uses the native AHV IPAM to inform the AWS DHCP server of all IP address assignments using API calls. NC2 relies on AWS to send gratuitous Address Resolution Protocol (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. For guest VMs, you can't share a given AWS subnet between two NC2 on AWS deployments. You can, however, use the same management subnet (AHV and CVMs) for multiple clusters.

The cloud network controller, a service in the AHV host, helps with ENI creation. The cloud network controller runs an OpenFlow controller, which manages the OVS in the AHV hosts and handles mapping, unmapping, and migrating guest-VM secondary IP addresses between ENIs or hosts. A subcomponent of the cloud network controller called cloud port manager provides the interface and manages AWS ENIs.

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

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



Overview of NC2 on AWS

By using native AWS networking, you quickly establish connectivity so that cloud administrators can focus on their tasks instead of managing additional networking technologies. The NC2 instance has full access to all AWS services, such as S3 and other EC2 instances running in the same Amazon Virtual Private Cloud. For a walkthrough of a typical deployment, see the Creating a Cluster section of the Nutanix Cloud Clusters on AWS Deployment and User Guide.

Flow Virtual Networking Flow Virtual Networking builds on the native networking stack in AWS to provide the same network virtualization and control that is offered in other NC2-supported clouds.

Flow Virtual Networking is a software-defined networking solution that provides multitenant isolation, self-service provisioning, and IP address preservation using Nutanix VPCs, subnets, and other virtual components that are separate from the physical network (AWS overlay) for the AHV clusters. It integrates tools to deploy networking features like virtual LANs (VLANs), VPCs, virtual private networks (VPNs), layer 2 virtual network extensions using a VPN or virtual tunnel end-points (VTEPs), and Border Gateway Protocol (BGP) sessions to support flexible networking that focuses on VMs and applications.

Flow Virtual Networking on AWS

Running Flow Virtual Networking requires that you run Prism Central on one of your NC2 on AWS clusters. You can have multiple NC2 instances consuming Flow Virtual Networking. The NC2 deployment process deploys Prism Central with high availability, and Prism Central hosts the control plane for Flow Virtual Networking.

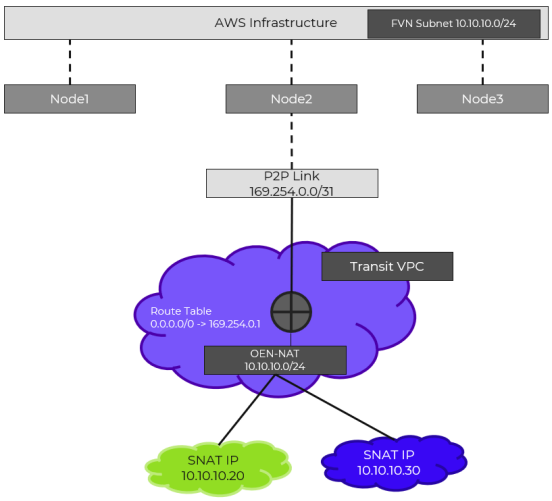
You need two new subnets when deploying Flow Virtual Networking: one for Prism Central and one for Flow Virtual Networking. The Prism Central subnet is automatically added to the Prism Element instance where it's deployed using a native AWS subnet. The Flow Virtual Networking subnet is also added to every node for each Prism Element instance that's managed by Prism Central using a native AWS subnet.

The Flow Virtual Networking subnet works as the external network for traffic from a Nutanix VPC. A VPC is an independent and isolated IP address space that functions as a logically isolated virtual network made of one or more subnets that are connected through a logical or virtual router. The IP addresses in a VPC must be unique, but IP addresses can overlap across VPCs.

In AWS, Nutanix uses a two-tier topology to provide external access to the Nutanix VPC. One of the NC2 hosts has a peer-to-peer link network setup that acts as an entry point outside the transit VPC. This peer-to-peer link network is a private network that doesn't overlap with any of your Nutanix user VPCs or the AWS environment.

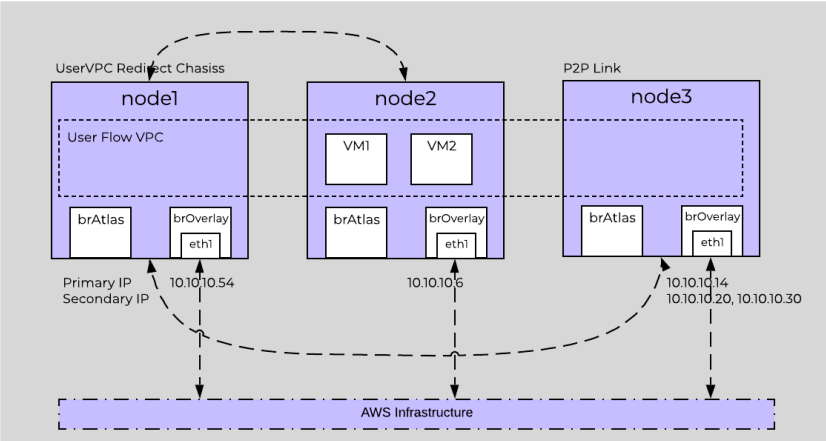
Deploying the first cluster in AWS with Flow Virtual Networking automatically creates a transit VPC that contains an external subnet called overlay-external-subnet-nat. The transit VPC requires this external subnet to send traffic to the peer-to-peer link. You can create an additional external network in the transit VPC for routed traffic, also known as no-NAT networking.

The Flow Virtual Networking native AWS subnet consumes the Source Network Address Translation (SNAT) IP addresses and any floating IP addresses that are given to user VMs that need inbound traffic to enter the user VPC. Each NC2 bare-metal node consumes the primary ENI IP address, and 60 percent of the native Flow Virtual Networking subnet range is available for floating IP addresses.



Path for Nutanix VPC Traffic to the AWS Network

Traffic exits the transit VPC through the peer-to-peer link that is hosted on one of the NC2 nodes. Even though all NC2 nodes have the Flow Virtual Networking subnet connected, the peer-to-peer link is only active on one host. Two VMs are located in the Nutanix VPC on node 2, and the UserVPC Redirect Chassis for those two VMs is located on node 1. After traffic exits node 1, it uses a Generic Network Virtualization Encapsulation (GENEVE) tunnel to exit the host with the peer-to-peer link.



Path for Nutanix VPC Traffic to the Peer-to-Peer Link

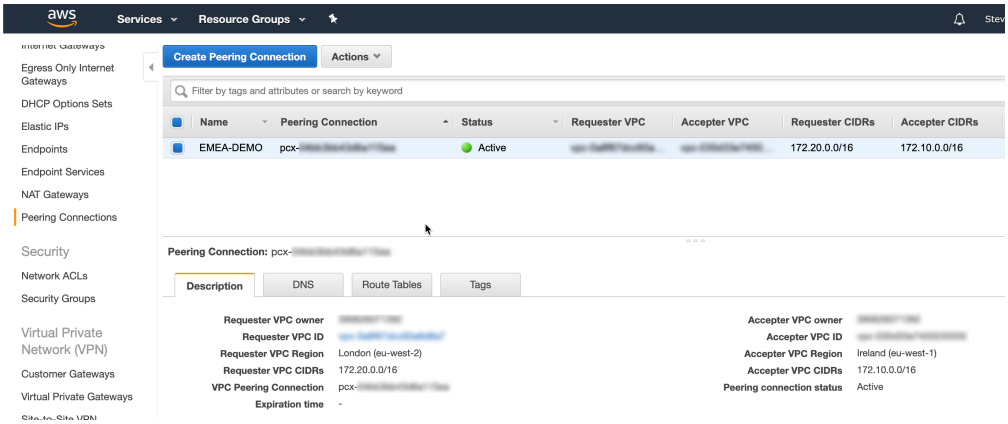
In the transit-VPC page in Prism Central, you can see which NC2 bare-metal node hosts the peer-to-peer link by clicking Associated External Subnets in the Summary tab.

WAN / L3 Networking

In most cases deployments will not be just in AWS and will need to communicate with the external world (Other VPCs, Internet or WAN).

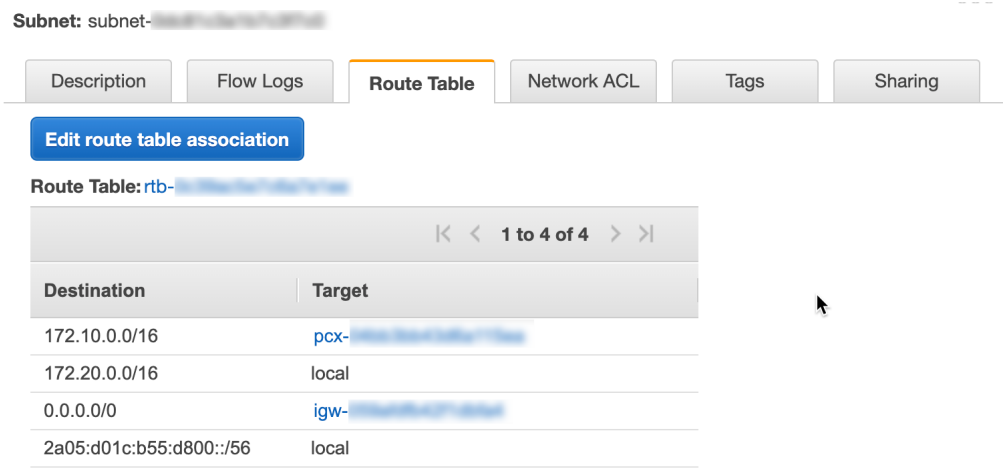
For connecting VPCs (in the same or different regions), you can use VPC peering which allows you to tunnel between VPCs. NOTE: you will need to ensure you follow WAN IP scheme best practices and there are no CIDR range overlaps between VPCs / subnets.

The following shows a VPC peering connection between a VPC in the eu-west-1 and eu-west-2 regions:



NC2A - VPC Peering

The route table for each VPC will then route traffic going to the other VPC over the peering connection (this will need to exist on both sides if communication needs to be bi-directional):



NC2A - Route Table

For network expansion to on-premises / WAN, either a VPN gateway (tunnel) or AWS Direct Connect can be leveraged.

Security

Given these resources are running in a cloud outside our full control security, data encryption and compliance is a very critical consideration.

The recommendations can be characterized with the following:

- Enable data encryption
- Only use private subnets (no public IP assignment)
- Lock down security groups and allowed ports / IP CIDR blocks
- For more granular security, leverage Flow

AWS Security Groups

You can use AWS security groups and network access control lists to secure your cluster relative to other AWS or on-premises resources. When you deploy Flow Virtual Networking, a fourth security group is deployed for Prism Central that has all the necessary rules for the Flow Virtual Networking control plane. If you plan to use Nutanix Disaster Recovery to protect AWS, edit this security group to allow traffic from the on-premises Prism Central instance. You can also use existing groups in your environment.

With AWS security groups, you can limit access to the AWS CVMs, AHV host, and guest VMs to only allow traffic from your on-premises management network and CVMs. You can control replication from on-premises to AWS down to the port level, and you can easily migrate workloads because the replication software is embedded in the CVMs at both ends. AOS 6.7 and later versions support custom AWS security groups, which provide additional flexibility so that AWS security groups can apply to the VPC domain and at the cluster and subnet levels.

Attaching the ENI to the bare-metal host applies your custom AWS security groups. You can use and reuse existing security groups across different clusters without additional scripting to maintain and support the prior custom security groups.

The cloud network service—a distributed service that runs in the CVM and provides cloud-specific back-end support for subnet management, IP address event handling, and security group management—uses tags to evaluate which security groups to attach to the network interfaces. You can use these tags with any AWS security group, including custom security groups. The following list is arranged in dependency order:

- **Scope**: VPC
  - **Key**: tag:nutanix:clusters:external
  - **Value**: <none> (leave this tag blank)

You can use this tag to protect multiple clusters in the same VPC.

- **Scope**: VPC or cluster
  - **Key**: tag:nutanix:clusters:external:cluster-uuid
  - **Value**: <cluster-uuid>

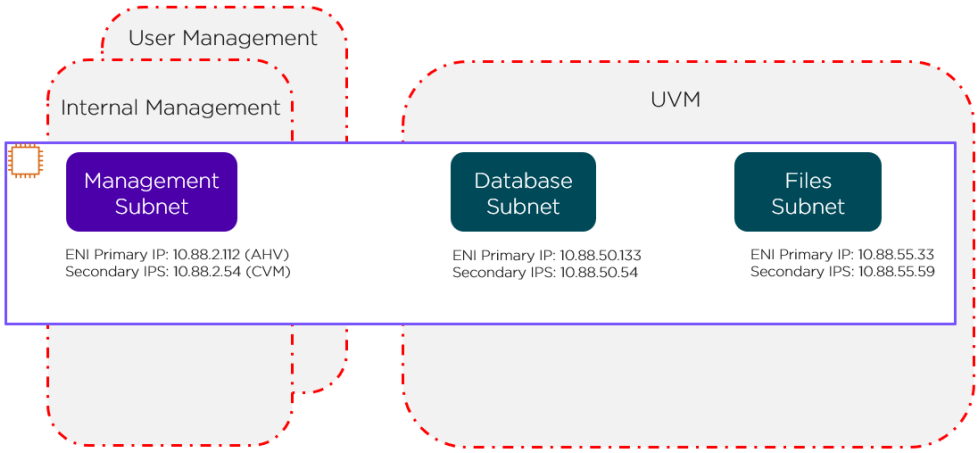
This tag protects all the UVMs and interfaces that the CVM and AHV use.

- **Scope:** VPC, cluster, network, or subnet
  - **Key:** tag:nutanix:clusters:external:networks
  - **Value:** <cidr1, cidr2, cidr3>

This tag only protects the subnets you provide.

If you want to apply a tag based on the subnet or CIDR, you need to set both external and cluster-uuid for the network or subnet tag to be applied. The following subsections provide configuration examples.

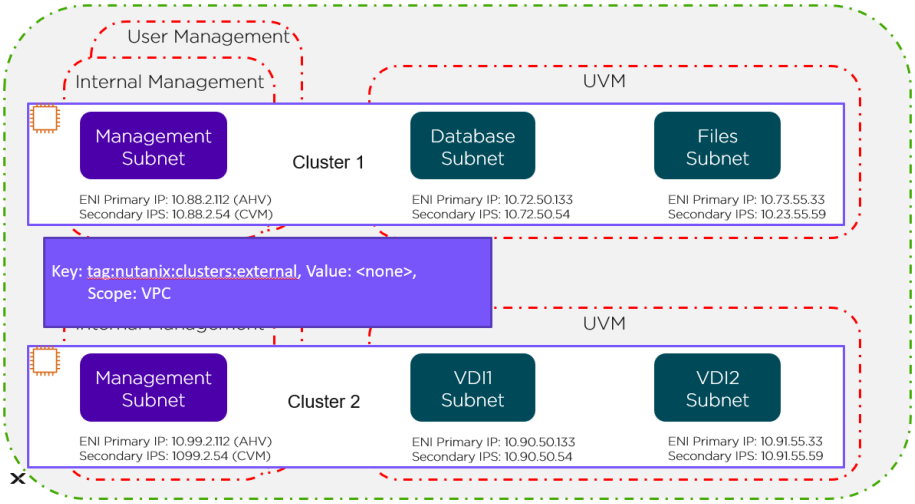
Default Security Groups



The red lines in the preceding figure represent the standard AWS Security Groups that deploy with the cluster.

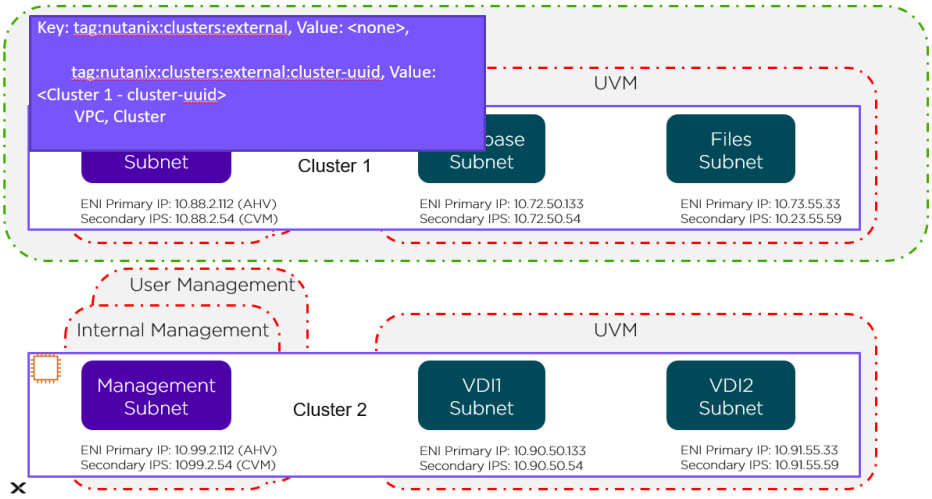
- Internal management Security Group: Allows all internal traffic between all CVMs and AHV hosts (EC2 bare-metal hosts). Don't edit this group without approval from Nutanix Support.
- User management Security Group: Allows users to access Prism Element and other services running on the CVM.
- UVM Security Group: Allows UVMs to talk to each other. By default, all UVMs on all subnets can talk to each other. This Security Group doesn't offer subnet granularity.

VPC Level



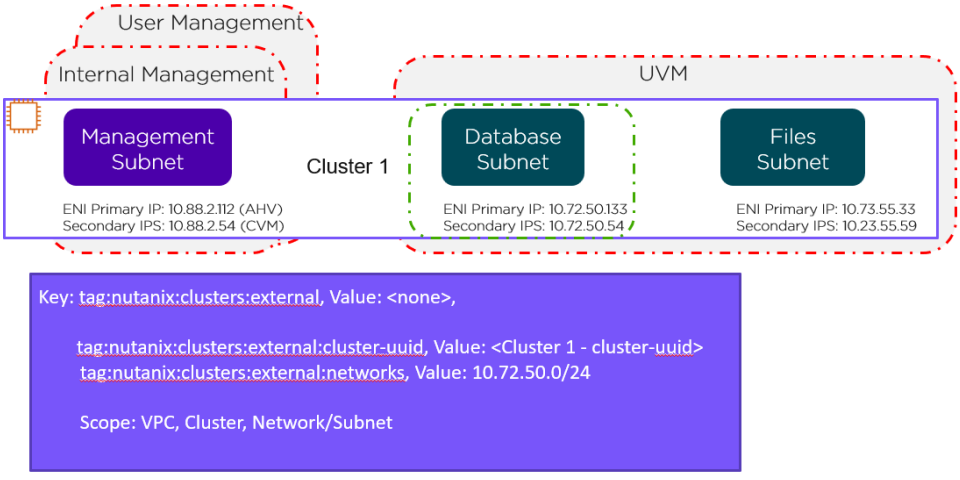
The green line in the preceding figure represents the VPC-level tag protecting Cluster 1 and Cluster 2.

Cluster Level



The green line in the preceding figure represents the cluster-level tag. Changes to these Security Groups affect the management subnet and all the UVMs running in Cluster 1.

Network Level



This network-level custom Security Group covers just the database subnet, as shown by the green line in the preceding figure. To cover the Files subnet with this Security Group, simply change the tag as follows:

- tag:nutanix:clusters:external:networks, Value: 10.72.50.0/24, 10.73.55.0/24

Usage and Configuration

The following sections cover how to configure and leverage NC2A.

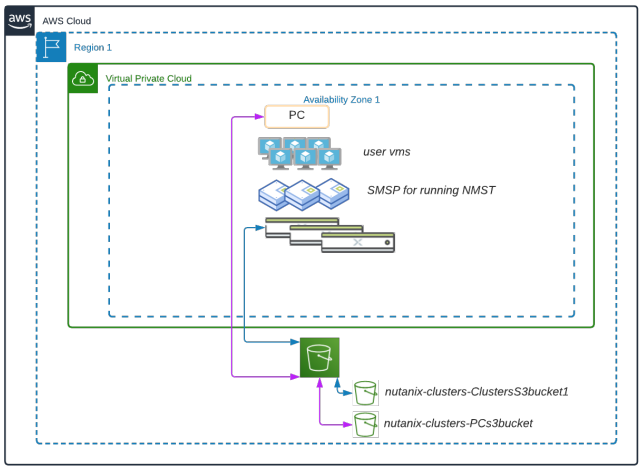
The high-level process can be characterized into the following high-level steps:

1. Create AWS Account(s)
2. Configure AWS network resources (if necessary)
3. Provision cluster(s) via Nutanix Clusters Portal
4. Leverage cluster resources once provisioning is complete
5. Protect your cluster

Native Backup with Nutanix Cluster Protection

Even when you migrate your application to the cloud, you still must provide all of the same day-two operations as you would if the application was on-premises. Nutanix Cluster Protection provides a native option for backing up Nutanix Cloud Clusters (NC2) running on AWS to S3 buckets, including user data and Prism Central with its configuration data. Cluster Protection backs up all user-created VMs and volume groups on the cluster.

As you migrate from on-premises to the cloud, you can be sure that there is another copy of your applications and data in the event of an AWS Availability Zone (AZ) failure. Nutanix already provides native protection for localized failures at the node and rack level, and Cluster Protection extends that protection to the cluster's AZ. Because this service is integrated, high-performance applications are minimally affected as the backup process uses native AOS snapshots to send the backup directly to S3.



Two Nutanix services help deliver Cluster Protection:

- Prism Central Disaster Recovery is responsible for backing up the Prism Central data. Instead of backing up to an AOS storage container, you can now supply a new S3 bucket to point the backup to.
- Nutanix Multicloud Snapshot Technology (NMST) replicates native Nutanix AOS snapshots to object storage. In the Cluster Protection design, the you supply a second new S3 bucket in AWS to send all the protected clusters' snapshots to the same S3 bucket. The Cloud Snapshot Engine runs on the Prism Central instance in AWS.

The following high-level process describes how to protect your clusters and Prism Central in AWS.

- Deploy a one- or three-node Prism Central instance in AWS.
- Create two S3 Buckets: one for Prism Central and one for your cloud clusters.
- Enable Prism Central protection.
- Deploy NMST.

- Protect your AWS cloud clusters.

The system takes both the Prism Central and AOS snapshots every hour and retains up to two snapshots in S3. A Nutanix Disaster Recovery category protects all of the user-created VMs and volume groups on the clusters. A service watches for create or delete events and assigns them a Cluster Protection category.

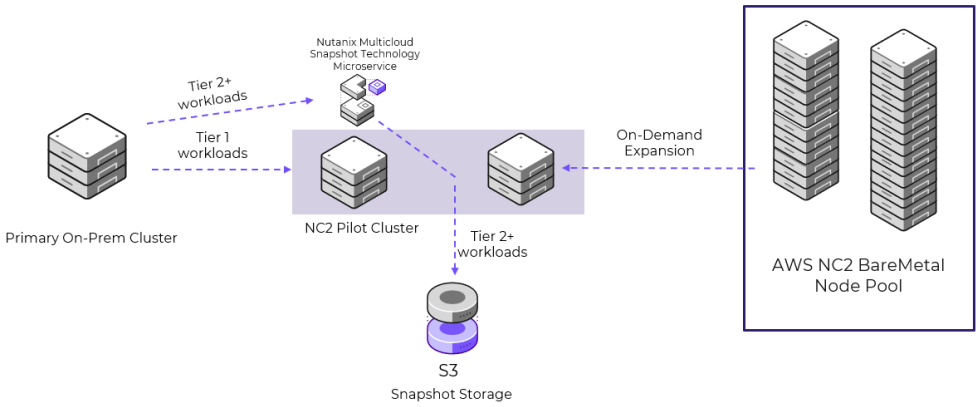
The following high-level process describes how to recover your Prism Central instances and clusters on AWS.

- The NC2 console automatically deploys a new NC2 cluster during the recovery process.
- Add your Prism Central subnet and any user VM networks to your recreated cloud cluster.
- Recover your Prism Central configuration from the S3 bucket.
- Register your Prism Central instance with the recovered cluster.
- Recover NMST.
- Create a recovery plan.
- Run the recovery plan from Prism Central.

Once the NMST is recovered, you can restore using the recovery plan in Prism Central. The recovery plan has all the VMs you need to restore. By using Nutanix Disaster Recovery with this new service, administrators can easily recover when disaster strikes.

Nutanix Disaster Recovery to S3 With NMST, you can send AOS snapshots from any Nutanix-based cluster to S3. This feature allows you to offload snapshots that you don't regularly access or applications that have higher recovery time objectives (RTOs) to optimize performance and capacity in the primary storage infrastructure.

When you use NMST on an NC2 cluster with at least three nodes (which we refer to as a pilot cluster), NMST redirects the AOS snapshots to S3. You can recover snapshots to the NC2 cluster or back to the primary site if a healthy Nutanix cluster is available. If you need additional space to recover the S3-based snapshots to the NC2 cluster, you can add more nodes to the cluster using the NC2 portal or the NC2 portal API.



This model supports fast RTOs for tier-1 applications by sending snapshots directly to the pilot cluster for quick restores and using S3 to store snapshots for tier-2 applications. Combined with using EBS as additional storage for tier-1 applications, this model can drastically reduce business costs while meeting requirements.