Book of Nutanix Cloud Clusters - Nutanix Cloud Clusters on AWS

[PDF generated August 17 2023. 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
- Etc.

Management interfaces(s):

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

Supported Environment(s):

- Cloud:
  - AWS
  - Azure

- 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.

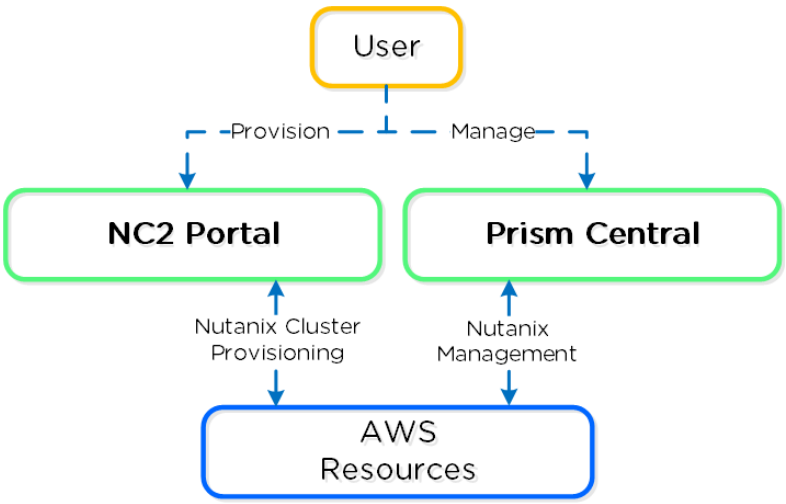
Cluster Architecture

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:

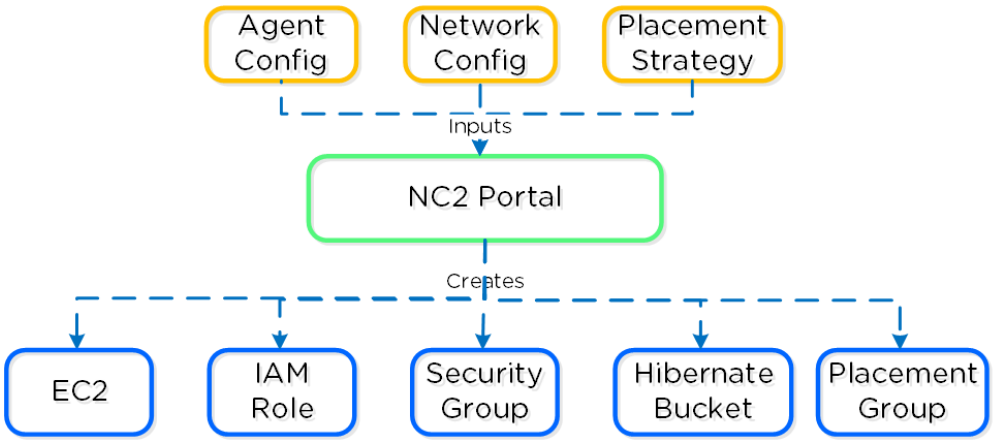
1. Create cluster in NC2 Portal
2. Deployment specific inputs (e.g. Region, AZ, Instance type, VPC/Subnets, etc.)
3. The NC2 Portal creates associated resources
4. Host agent in Nutanix AMI checks-in with Nutanix Clusters on AWS
5. Once all hosts as up, cluster is created

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



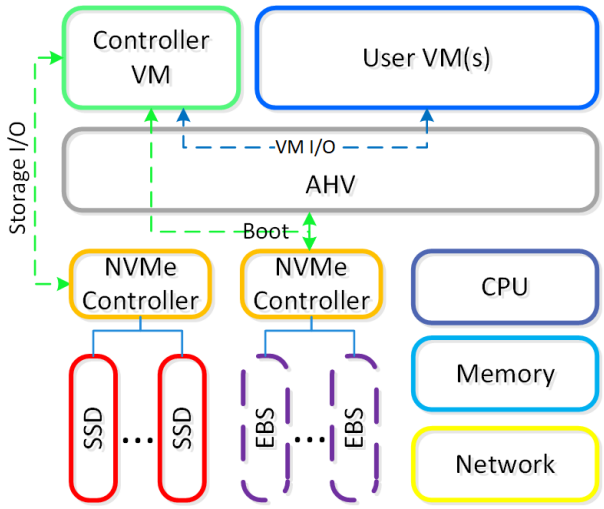
NC2A - Overview

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

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.

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.

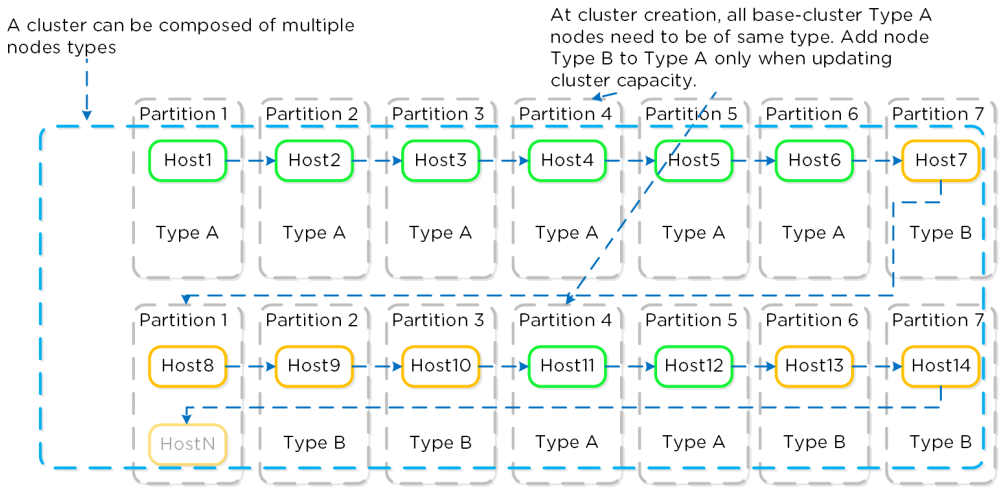
NC2 on AWS supports combining heterogenous node types in a cluster. You can deploy a cluster of one node type and then expand that cluster's capacity by adding heterogenous nodes to it. This feature protects your cluster if its original node type runs out in the Region and provides flexibility when expanding 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.

When combining instance types in a cluster, you must always maintain at least three nodes of the original type you deployed the base cluster with. You can expand or shrink the base cluster with any number of heterogenous nodes if at least three nodes of the original type remain and the cluster size stays within the limit of 28 nodes.

The following table and figure both refer to the cluster's original instance type as Type A and its compatible heterogenous type as Type B.

Table: Supported Instance Type Combinations

Type A	Type B
i3.metal	i3en.metal
i3en.metal	i3.metal
z1d.metal	m5d.metal
m5d.metal	z1d.metal



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

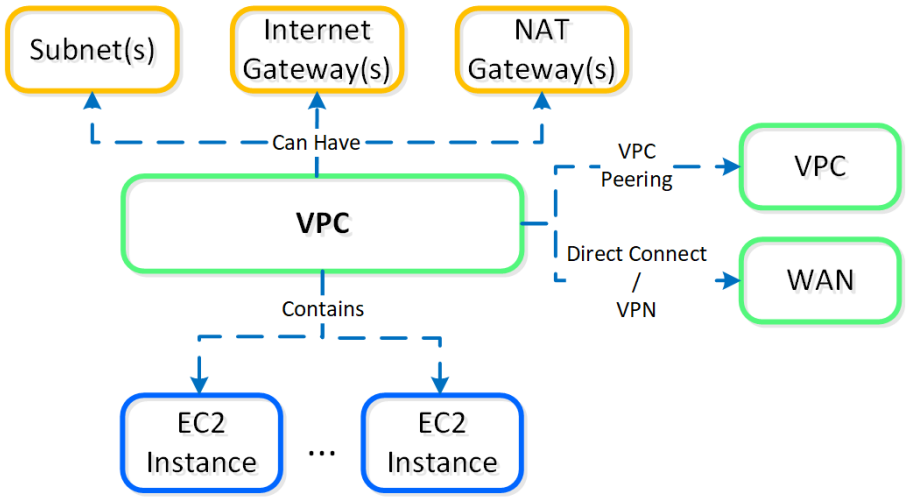
Networking can be broken down into a few core areas:

- 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

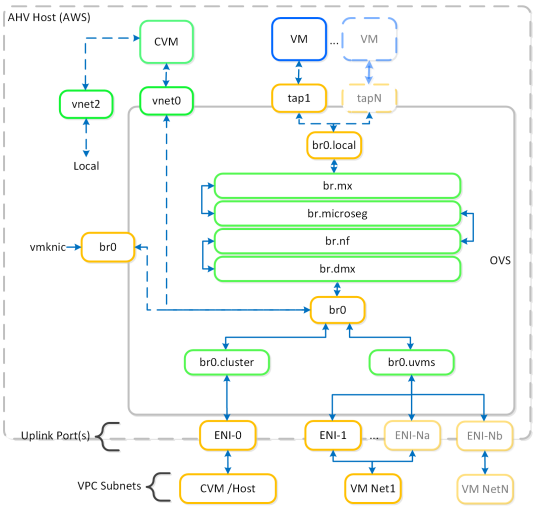
AWS will create a default VPC/Subnet/Etc. with a 172.31.0.0/16 ip scheme for each region.

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.

Host Networking

The hosts running on baremetal in AWS are traditional AHV hosts, and thus leverage the same OVS based network stack.

The following shows a high-level overview of a AWS AHV host's OVS stack:

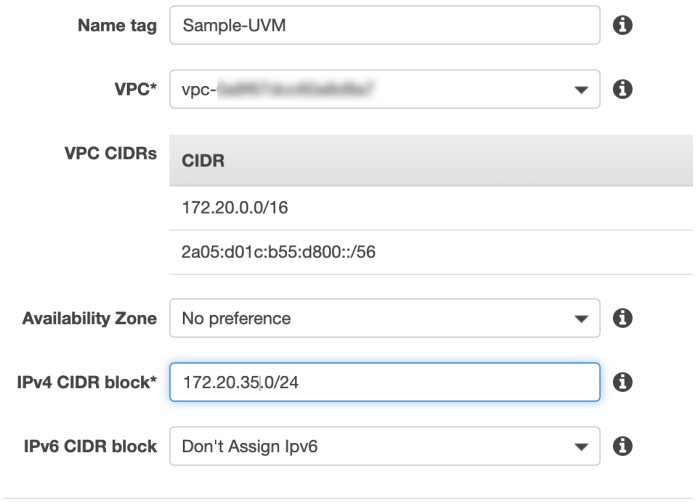


NC2A - OVS Architecture

The OVS stack is relatively the same as any AHV host except for the addition of the L3 uplink bridge.

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:

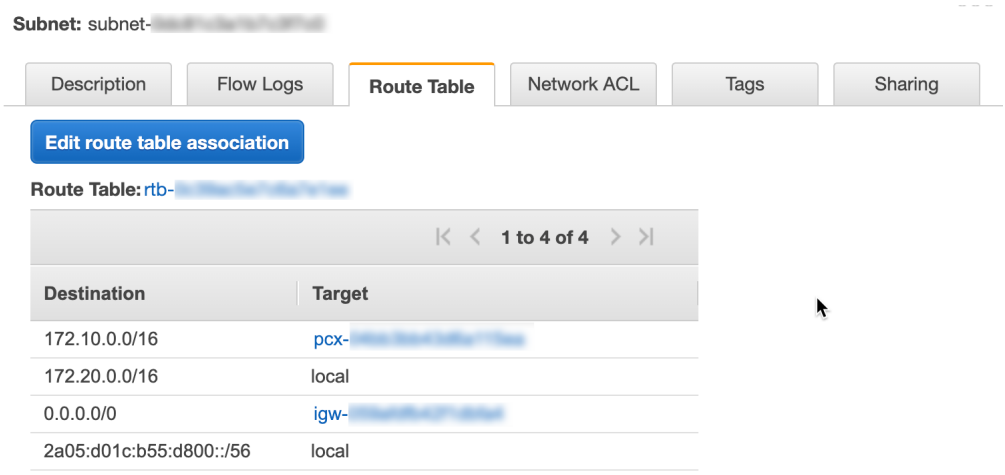
From the AWS VPC dashboard, click on 'subnets' then click on 'Create Subnet' and input the network details:



NC2A - OVS Architecture

NOTE: the CIDR block should be a subset of the VPC CIDR range.

The subnet will inherit the route table from the VPC:



NC2A - Route Table

In this case you can see any traffic in the peered VPC will go over the VPC peering link and any external traffic will go over the internet gateway.

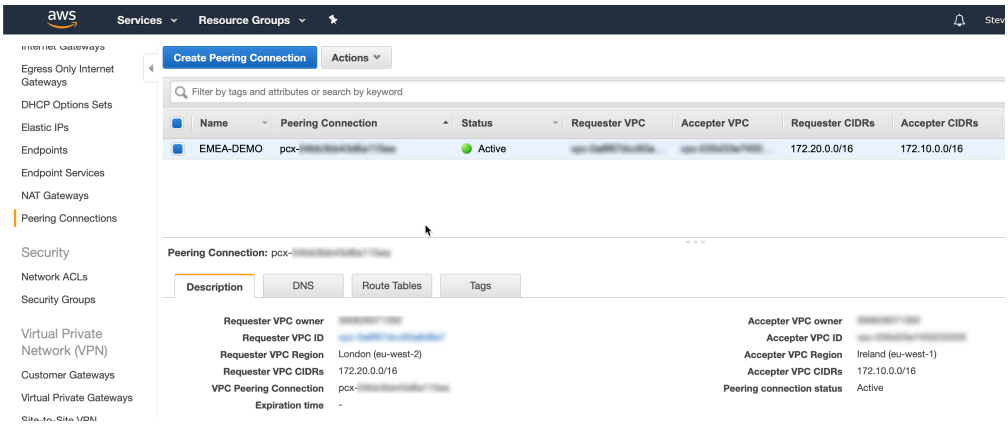
Once complete, you will see the network is available in Prism.

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).

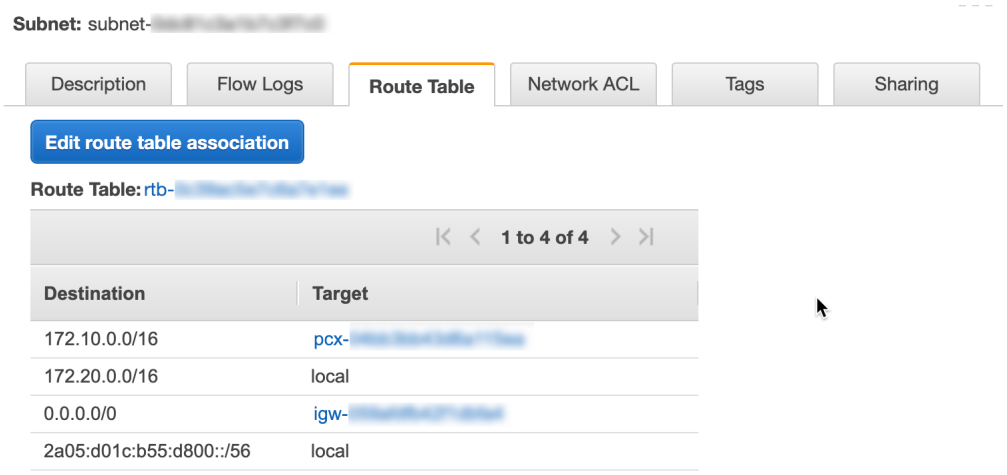
For connecting VPCs (in the same or different regions), you can use VPC peering which allows you to tunnel between VPCs. NOTE: you will need to ensure you follow WAN IP scheme best practices and there are no CIDR range overlaps between VPCs / subnets.

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

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

More to come!