v1.0 | August 2022 | RA-2147

REFERENCE ARCHITECTURE

Nutanix for Enterprise Edge Computing

NUTANIX
YOUR ENTERPRISE CLOUD

Copyright

Contents

1. Executive Summary

Each organization defines edge computing differently; we define edge computing as workloads (applications and services) hosted and contained outside the core enterprise datacenter. These workloads fall into three categories:

1. Internet of Things (IoT) edge
   a. End users: developers, data scientists
   b. Roles: data acquisition, inferencing, modeling
   c. Typical use cases: manufacturing, transportation, robotics
2. Enterprise edge
   a. End user: enterprise IT
   b. Roles: localized applications, service hosting, remote office
   c. Typical use cases: file server, remote desktops, latency sensitive applications, business continuity planning (BCP)
3. Telco edge
   a. End user: the enterprise (operated and managed by the service provider)
   b. Typical use cases: multi-access edge computing, carrier-grade network function virtualization (NFV) services

This reference architecture focuses on the Nutanix Cloud Platform as a solution for enterprise edge workloads.

2. Introduction

Purpose

We created this reference architecture to assist architects through the design process of implementing enterprise edge sites. After reading this document, architects should understand Nutanix best practices for enterprise edge sites and the implications and justifications of the design choices they make.

Audience

This reference architecture is part of the Nutanix Solutions Library. We wrote it for IT decision makers and architects responsible for deploying enterprise edge applications and services. Readers of this document should already be familiar with Nutanix.

This document often refers readers to the Nutanix Hybrid Cloud Reference Architecture.

Unless otherwise stated, the solution described in this document is valid on AOS versions 5.20 and later.

Document Version History

Version Number	Published	Notes
1.0	July 2022	Original publication.

3. Nutanix Cloud Platform Overview

Physical Layer

Nutanix converges compute, storage, storage networking, and virtualization, reducing complexity and replacing the separate servers, storage systems, and storage area networks (SANs) found in conventional datacenter architectures. Each node in a Nutanix cluster includes compute, memory, and storage, and nodes are grouped into clusters. AOS software running on each node pools storage across nodes and distributes operating functions across all nodes in the cluster for performance, scalability, and resilience.

Hardware

The Nutanix Cloud Platform provides significant flexibility when it comes to hardware platform selection. Available options include:

- Nutanix NX appliances
- Original equipment manufacturer (OEM) appliances from leading vendors such as Dell, Lenovo, HPE, and Fujitsu
- Other third-party servers from a wide range of vendors

The Nutanix Hardware Compatibility Lists contain the most up-to-date information on supported systems.

Compute

The process for sizing systems to meet compute needs in a Nutanix environment is similar to that of other architectures. However, you must ensure that your design provides enough compute (CPU and RAM) to support the CVM.

Storage

Nutanix nodes offer a range of storage configurations:

- Hybrid nodes combine flash SSDs for performance and HDDs for capacity.
- All-flash nodes use traditional flash SSDs.
- NVMe nodes use NVMe SSDs.

You can mix different node types in the same cluster.

Nutanix AOS storage exhibits data avoidance and efficiency using techniques such as thin provisioning, intelligent cloning, compression, deduplication, and erasure coding. These techniques accelerate application performance and optimize storage capacity. They are intelligent and adaptive, requiring little or no fine-tuning in most cases, which reduces operating expenses and frees your IT staff to focus on growth and innovation. Unlike traditional storage architectures, the Nutanix web-scale design ensures that data efficiency techniques scale as the cluster grows.

For more information, see the Data Efficiency tech note.

Networking

The distributed storage architecture relies on the performance and resilience of the physical network. A good design provides high performance while maintaining simplicity. For more details on networking, see the Physical Networking best practice guide and the AHV Networking best practice guide.

Cluster Design

A Nutanix cluster is the management boundary of the storage provided to a group of workloads. A Nutanix deployment can have either single clusters with mixed workloads or dedicated clusters for each workload type in a block-and-pod design. Designs with dedicated clusters can include the following cluster types:

Management clusters

Designed to run VMs that support datacenter management services, such as Nutanix Prism Central, VMware vCenter, and Active Directory (AD) domain controllers, and other management workloads, such as DNS, DHCP, NTP, and syslog.

Edge clusters

Reside at an edge or remote and branch office (ROBO) deployment to run VMs or ROBO workloads. These clusters are typically distinguished from normal workload clusters by their small size and limited external bandwidth and can contain a variety of node types and sizes.

Hypervisors

This solution covers two enterprise-grade hypervisors: Nutanix AHV and VMware ESXi (vSphere). Both fulfill a similar set of requirements and use cases. Nutanix AHV is included at no additional cost with every Nutanix node. For more information on AHV, see the Nutanix AHV Virtualization tech note and the AHV best practice guide. For information on VMware vSphere on Nutanix, see the VMware vSphere best practice guide.

Note: Nutanix also supports Microsoft Hyper-V; however, this document only covers AHV and ESXi.

Management

Nutanix Prism is the centralized management solution for Nutanix environments. Prism combines multiple aspects of datacenter management into a single consumer-grade design that provides complete infrastructure and virtualization management, operational insights, and troubleshooting.

For more detailed information about Prism, see the Prism software documentation and the Prism tech note.

Business Continuity and Disaster Recovery

We built Nutanix with resilience in mind, so the platform includes redundancy for power and other hardware components and enables you to protect entire datacenters. Nutanix provides native business continuity, including backup and restore and disaster recovery, at the hardware, node, data, VM, and datacenter or site levels.

For more detailed information about Nutanix business continuity and disaster recovery (BCDR), read the Data Protection and Disaster Recovery best practices guide, the Disaster Recovery with Nutanix Cloud Clusters (NC2) on AWS tech note, the Nutanix Disaster Recovery (previously Leap) software documentation, and the Nutanix Mine software documentation.

Snapshots and Clones

An individual cluster can create a snapshot of a vDisk or an entire VM. Nutanix snapshots are redirect-on-write, where the pointers that indicate which physical disk extent maps to a virtual disk are updated only as data changes.

Snapshots can be application-consistent or crash-consistent. AOS can trigger an application-consistent snapshot in Nutanix Guest Tools or VMware tools, depending on which hypervisor you use.

For ESXi-triggered clones, AOS supports offloading the clone using vStorage API for Array Integration (VAAI). The VM must be turned off and the clone must reside on the same datastore as the primary VM in order to trigger offloading.

> Note: Nutanix snapshots are a vital element of a data protection strategy; however, they aren't a substitute for a full backup methodology.

Replication Options

You can also replicate Nutanix snapshots to multiple destination clusters to provide multiple redundant copies if needed. Configure these arrangements as part of the snapshot schedule by specifying each remote site that should receive a copy and how many copies it should retain.

> Note: One-to-many relationships are allowed (Site A to Site B, Site A to Site C, Site A to Site N), but cascading relationships (Site A to Site B to Site N) aren't supported.

Protection Domains

With Prism Element, you can group VMs into protection domains. During a failover event, all the entities in the protection domain are activated at the remote site as a group. The RPO you choose for these entities determines whether the system uses traditional asynchronous snapshots or lightweight snapshots for this process.

You can also create a consistency group for all VMs and volume groups in a protection domain. Consistency groups enable you to snapshot all members of the group in a crash-consistent manner.

Automation

Automation and orchestration are critical to IT success. By simplifying infrastructure management across the entire lifecycle, automating operations, and enabling self-service, Nutanix helps you deploy datacenter infrastructure that delivers a high degree of scalability, availability, and flexibility with Prism (centralized automation management), Nutanix Cloud Manager (NCM) Self-Service (previously Calm; application orchestration), and simplified test and development automation.

Security and Compliance

Nutanix takes a security-first approach that includes a secure platform, extensive automation, and a robust partner ecosystem. Security Configuration Management Automation (SCMA) monitors the health of storage and VMs, automatically healing any deviations from the baseline, and data-at-rest encryption (DaRE) features add to our robust security capabilities. We also provide additional configuration options if you need to add an extra layer of security to fulfill business or technical requirements.

For more detailed information on Nutanix security and compliance features, see the following documents, separated by category:

- Networking: Flow Network Security tech note
- Information: Information Security tech note
- Cloud: Nutanix DRaaS Security tech note
- Database: Nutanix Era Security tech note

4. High-Level Design Considerations

There are several high-level design decisions you must make before proceeding to the detailed technical design.

Choosing an Enterprise Edge Architecture

First, you need to know what availability you need and what failures the design must protect against. These are key input requirements.

You also need to be familiar with the regions and availability zones as defined in the High-Level Design Considerations section of the Nutanix Hybrid Cloud Reference Architecture. Briefly, a region holds one or more availability zones. Each availability zone contains one or more datacenters or edge locations.

Figure 1: Availability Zones in a Region

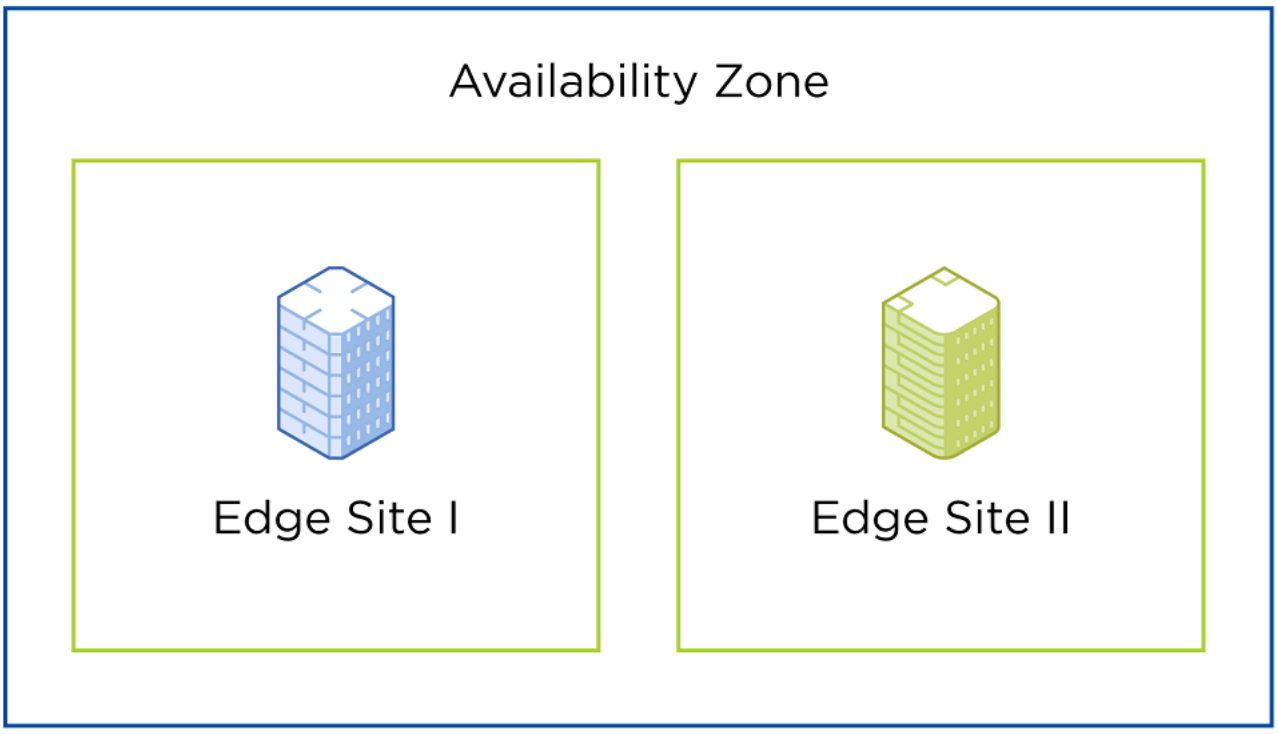

Figure 2: Edge Sites in an Availability Zone

Because an enterprise edge deployment can cover many locations often distributed across multiple geographies, we suggest that you logically group the manageability, replication, and BCDR components of your edge sites into a single region with multiple availability zones. This logical grouping provides design advantages for the control plane, scalability and performance, and BCDR.

> Note: If your enterprise edge deployment has a small number of sites and single locations, you may not need this grouping.

When choosing an enterprise edge architecture, keep the following things in mind:

- Latency requirements can vary depending on the deployment and the resources being consumed. For example, witness VMs hosted in the cloud can tolerate up to 500 ms of latency per 100 instances, while Prism Central can only tolerate 150 ms of latency per Prism Element instance that it manages.

- Bandwidth requirements vary based on desired recovery point objectives (RPOs), replication mediums, and the resources being consumed.
- A scaled-out Prism Central instance can manage 400 clusters and up to 2,000 nodes.
- You need accurate NTP access.
- You must consider whether you want a WAN-restricted topology or open internet access.
- You need to review your BCDR requirements, planned versus unplanned downtime, and environmental risk.
- You can replicate data to more than one location using different replication types if needed. For example, you might use NearSync replication between sites located in the same availability zone and asynchronous replication to a datacenter in another region.

Edge Operating Models

Nutanix supports the following enterprise edge operating models, independent of the operating model of the core datacenter:

- Single site, active
  - Applications are active in the edge site.
  - Backups are stored in the same edge site that the applications run in.
  - Long-time archiving is stored offsite.
  - This model provides limited protection against datacenter failure.
- Single site, multiple failure domains, active-active
  - Applications are active in one or more failure domains during normal production.
  - Failure domains can be in different locations in the site (for example, opposite ends of a campus).
  - Failure domains can provide disaster avoidance and disaster recovery for each other.

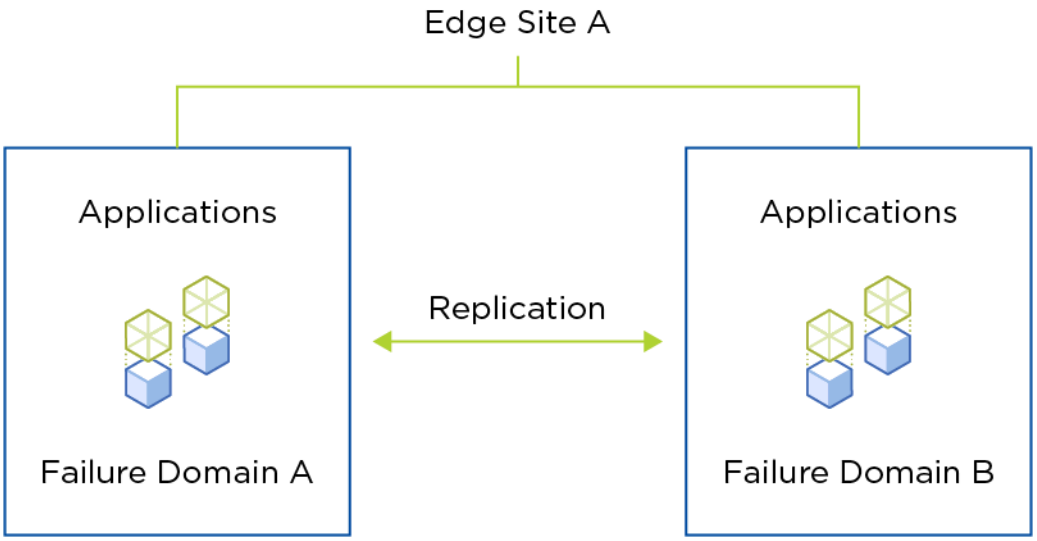

Figure 3: Single Site, Multiple Availability Domains, Active-Active

- Multiple sites, active-passive
  - Edge-based applications typically run at the edge site during normal production.
  - Usually the central datacenter (passive site) provides disaster avoidance and disaster recovery for the active enterprise edge site.

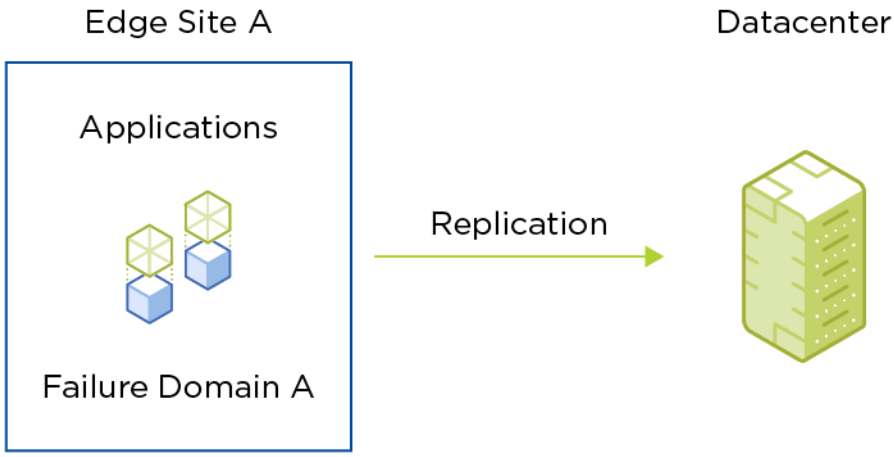

Figure 4: Single Site, Multiple Availability Domains, Active-Passive

- Multiple sites, fan-in or central datacenter
  - Multiple enterprise edge sites replicate data to the datacenter.
  - The datacenter acts as disaster recovery for the edge sites.

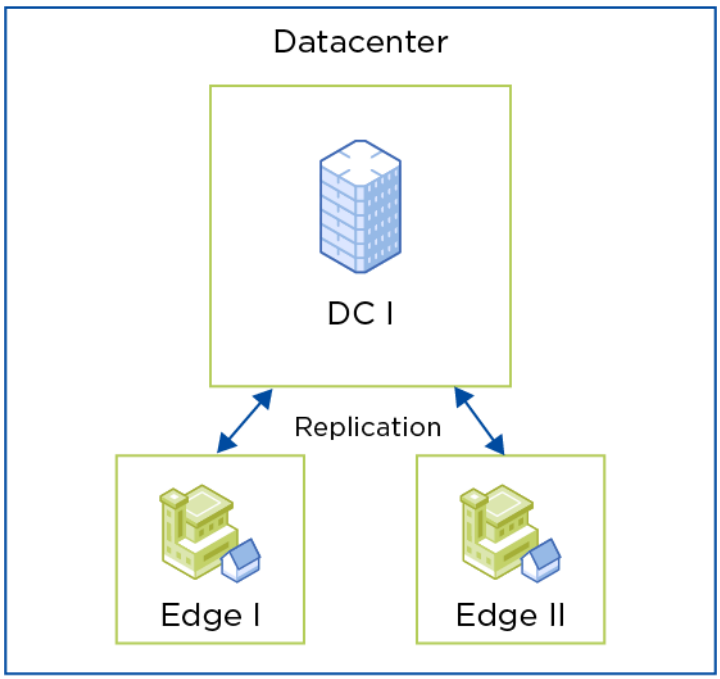

Figure 5: Multiple Sites, Fan-In or Central Datacenter

- Multiple sites, chain structure
  - The datacenter provides disaster recovery for the closest edge site or sites.
  - This model is often used by organizations with multiple datacenters or regions.
  - The first-tier edge sites connected to the datacenter acts as disaster recovery sites for the next tier of edge sites. Some organizations call these locations regional datacenters.

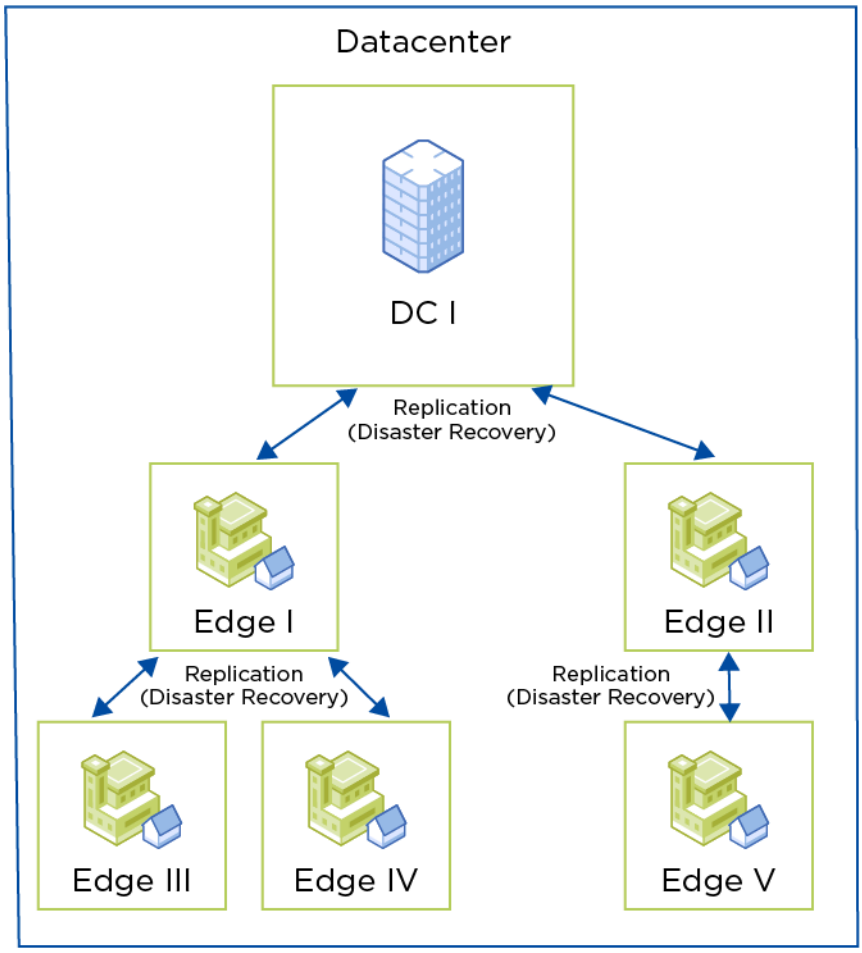

Figure 6: Multiple Sites, Chain Structure

Choosing a Hypervisor

Nutanix supports multiple hypervisors: Nutanix AHV, VMware vSphere, and Microsoft Hyper-V. This design covers the deployment of either Nutanix AHV or VMware vSphere.

Nutanix AHV is included with AOS and delivers high performance, flexible migrations, integrated networking, security hardening, automated data protection and disaster recovery, and rich analytics. AHV is a lean virtualization solution with robust, integrated management features. VMware vSphere is a proven virtualization platform used by many organizations. It has a robust

ecosystem but is a complex platform with many design choices and settings to tune that often requires the purchase of additional licenses.

System Availability

When evaluating availability requirements, start by building a dependency map of the applications and services hosted at the enterprise edge coupled with the service-level agreement (SLA; measured in 9s of availability) that the Nutanix system needs to support. Next, categorize availability events as planned or unplanned to enable you to plan for unforeseen events.

Single-Node Clusters

Let's say, for example, that an edge site has a single node in the Nutanix cluster (ROBO1) that hosts compute and storage. During write operations, the cluster mirrors data across the internal drives. In the event of a drive loss, the cluster enters a read-only state until the drive failure has been addressed, meaning that any hardware failure renders the system, application, and service offline.

You could address the availability constraints of the single-node cluster in the following ways:

Replicate data to a neighboring node: Deploy and maintain two separate single-node clusters that each serve as a replication partner to the other and independent failure domains.

Replicate data to the central datacenter or cloud.

The advantages of the single-node cluster are:

- It's an operationally well-understood and simplified model.
- It has lower capex.
- It offers 1 GbE support.

The disadvantages of the single-node cluster are:

- It can support fewer VMs.
- You can't easily expand the cluster; it requires additional planning.
- It doesn't support storage optimization features.

- Without standard operating procedures (SOPs), RPOs and RTOs can be negatively affected.
- Planned outages such as patches and updates can require downtime.

Two-Node Clusters

A two-node cluster (ROBO2) improves system availability. In this design, two Nutanix nodes replicate data synchronously (RPO 0) while supporting VM high availability and migration features like automatic restart and migration, dramatically decreasing the RTO.

Two-node clusters offer reliability for smaller sites that must be cost effective and run with tight margins. These clusters only use a witness in failure scenarios to coordinate rebuilding data and automatic upgrades. You can deploy the witness offsite up to 500 ms away for asynchronous replication and 200 ms away for Metro Availability. Multiple clusters can use the same witness for two-node and metro clusters.

The advantages of the two-node cluster are:

- It aggregates system resources (compute and storage).
- It reduces the RPO and RTO.
- It automatically recovers resources.
- There's minimal impact from planned outages such as patches and updates.
- It supports asynchronous replication with lower RPOs to a neighboring cluster.

The disadvantages of the two-node cluster are:

- You can't easily expand the cluster; it requires additional planning.
- It relies heavily on the availability of a witness VM.
- Staff needs familiarity with SOPs for node patching and upgrading.
- It doesn't support storage optimization features.

Three-Node or Greater Clusters

Clusters with three or more nodes (ROBO3) follow the traditional datacenter deployment strategy and have the same levels of availability and platform-specific features to optimize compute and storage resources. Clusters with three or more nodes address availability concerns by providing advanced clustering and data distribution, which result in lower RTOs for planned versus unplanned system outages.

The advantages of clusters with three or more nodes are:

- They support all Nutanix features and functionalities.
- Their clustering and data distribution yield a lower RTO.
- They have zero down time for patches, upgrades, and cluster expansion operations.

The disadvantages of clusters with three or more nodes are:

- They have a higher capex due to the additional nodes.
- For optimum performance, they require a 10 GbE network connection for all nodes. However, 1 GbE is sufficient for clusters with up to eight members.
- They use additional compute resources.
- They have increased power and cooling requirements.

System Recoverability

System recoverability defines how you recover the service or application hosted at the edge when the primary copy has been compromised or lost.

Nutanix has several recoverability options for customers who want to protect and back up their data against defined SLAs. The following figure and table illustrate four common recoverability strategies that you can use regardless of the edge strategy you deployed.

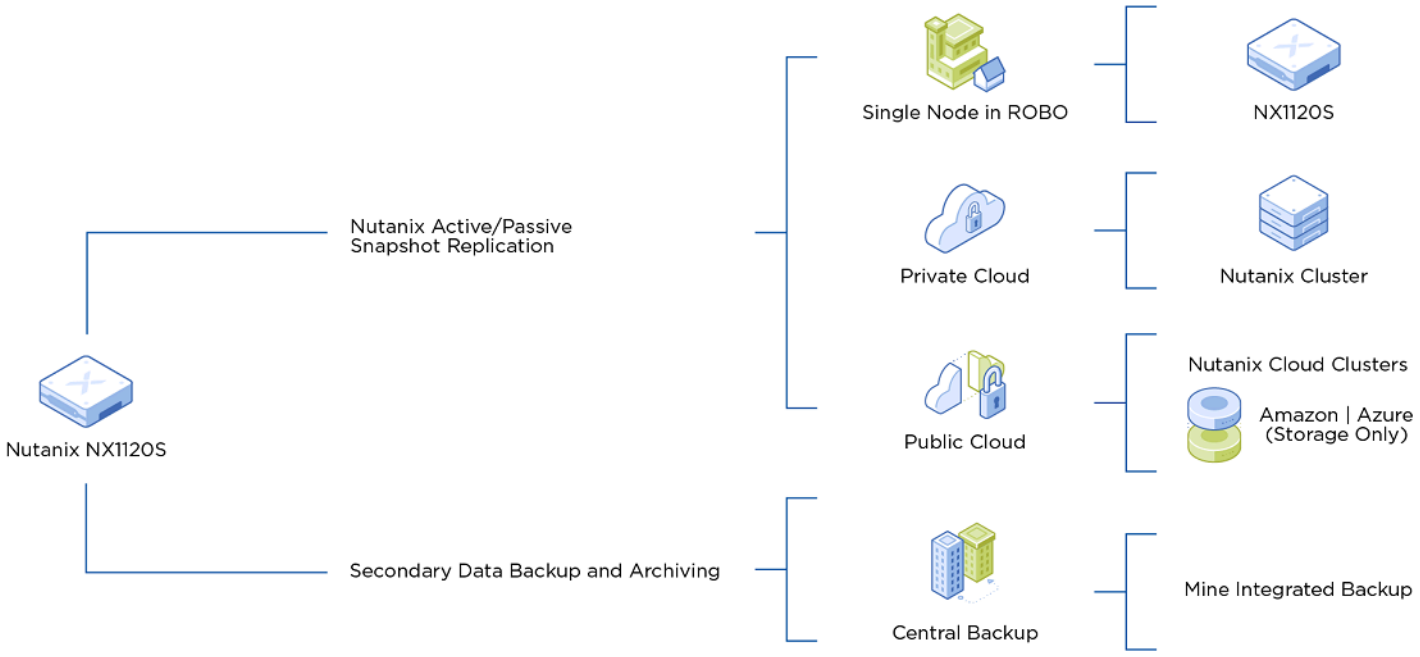

Figure 7: Overview of Nutanix Recoverability Options

Table: Replication Options with Nutanix

Destination	Strategy	Description	RPO	Recovery Method
Single-node ROBO	Asynchronous replication	Used when the source and the destination are single-node clusters.	6 hours	VM
Private cloud	Near-synchronous (NearSync) replication	Used when there are at least three nodes in a cluster on both the primary and the remote site.	1–15 minutes	VM

Destination	Strategy	Description	RPO	Recovery Method
Private cloud	Asynchronous replication	Used when replicating to a cluster of at least three nodes in the primary datacenter.	60 minutes	VM
Public cloud	Nutanix Cloud Clusters (NC2; asynchronous)	Used when customers need to host Nutanix on AWS, GCP, or the Nutanix public cloud; allows customers to use the same replication strategies as the private cloud.	60 minutes	VM
Public cloud	Amazon S3 / EBS, Azure Blob	Uses the Cloud Connect feature to replicate VMs to AWS S3 (EBS for metadata) or Azure Blob Storage. With this option, customers can't turn on VMs because they're just using a storage target or service in the public cloud.	60 minutes	VM

Destination	Strategy	Description	RPO	Recovery Method
Central backup	Nutanix Mine (secondary data backup and archiving)	Used when customers need a dedicated, turnkey solution that supports backup and archiving using integrated solutions from Veeam and HYCU.	60 minutes	VM- or file-level backup

For more information on NearSync requirements and limitations, see Requirements of Data Protection with NearSync Replication. For more information on single-node clusters, consult Single-node Clusters. To understand more about the Cloud Connect feature and protection domains, see Asynchronous Replication Using Cloud Connect (On-Prem to Cloud).

Performance and Scalability

When you build an enterprise edge solution, you need to carefully define your performance and scalability requirements because node deployment strategy can adversely affect the scalability of the solution. For example, with ROBO1, you're limited to a single node and can't expand your deployment by adding additional nodes or take advantage of data distribution across multiple compute nodes. ROBO2 has similar constraints.

> Note: If you need to be able to scale your edge solution beyond two nodes or add storage-only nodes to the cluster, you should implement a three-node cluster.

Regardless of whether your sites are classified as small (fewer than 10 VMs) or large (more than 50 VMs), you need to standardize the resources the edge sites provide and consume from the core datacenter. To accomplish this standardization, consider the four key elements of edge sites:

1. Devices and applications
   a. Create a dependency map of physical devices and applications that your infrastructure interfaces with or provides services for.
   b. Understand whether you plan to scale your devices and applications linearly over time.
   c. Establish whether the device or application should have a dedicated cluster or share a mixed cluster with other applications and services.
2. Clusters
   a. Define the maximum size of the clusters in your edge sites and the maximum number of edge sites in your organization.
   b. Reference the configuration maximums of Nutanix components and third-party components used in your design.
   c. Define the services that edge sites should provide (file sharing, NFV).
3. Network
   a. Choose the optimal network layout, considering how sites receive and send public versus private traffic and connect to the core datacenter.
   b. Use minimal bandwidth and a latency standardized for connectivity to the core datacenter.
4. Cloud
   a. Determine your cloud deployment model (hybrid, private, public).
   b. Determine which additional cloud resources you need, such as additional storage to cover replication from additional edge sites.
   c. Determine how much additional bandwidth you need in order to add additional edge sites to the cloud.

Manageability

The deployment model you use for the enterprise edge solution influences overall manageability.

Dedicated or Mixed Clusters

Typically, in enterprise edge sites organizations deploy between 5 and 10 VMs. In this instance, a mixed workload cluster makes the most sense. However,

with larger deployments (for example, large distribution centers) that serve many mission-critical applications and have many regulatory compliance requirements, the complexity of sizing and operating the mixed environment increases dramatically.

When you operate edge sites at a large scale, your architecture, design choices, and requirements start to look more like those of a traditional datacenter. For these cases, we encourage you to review the Will You Deploy Dedicated or Mixed Clusters? section of the Nutanix Hybrid Cloud Reference Architecture.

Storage-Only Nodes in Clusters

Storage-only nodes add storage capacity and I/O performance to a cluster. Storage-only nodes can be any node type, but they're typically configured with just enough CPU and memory resources to run the CVM because no application VMs run on them. These nodes are members of the AOS cluster but aren't visible to the hypervisor cluster for nonstorage functions, so the hypervisor can't schedule other VMs to run on them.

Note: You can only add storage-only nodes to clusters with three or more nodes.

Note: We recommend deploying storage-only nodes in pairs.

Choosing a Control Plane

When working with larger edge deployments geographically dispersed across multiple regions, we expect the organization to face control plane proliferation, so you must give additional thought to the operational administration of the clusters. In this design:

- The primary control plane is Prism Central.
- If you use ESXi, you also need VMware vCenter.

Each of these control planes has a maximum size that dictates the number of VMs, nodes, and clusters a single instance can manage. For AOS 5.20 (the minimum version required for these solutions), Prism Central (scaled-out configuration) can manage a maximum of:

- 25,000 VMs

- 400 clusters
- 2,000 nodes

These limits are quickly realized when you work with multiple sites and auxiliary solutions, especially when sharing a Prism Central instance with the core datacenter infrastructure. At very large scale (multiple sites, geographically dispersed), it makes sense to deploy an additional Prism Central instance in the management cluster in your core datacenter to manage the edge sites or to deploy a Prism Central instance specific to the region it manages.

Choosing the Right Licensing and Support

When evaluating the different Nutanix platform options available, there are two key decision points:

- Your software and support licensing types
- Your platform vendor

Software and Support Licensing

Nutanix nodes are primarily available in two different purchasing or licensing options: appliances or software-only.

Appliances:

- Available directly from Nutanix or through our OEM relationships.
- Appliance-based licensing is referred to as life-of-device licensing, meaning it's only applicable to the appliance it was purchased with.
- The manufacturer of the appliance takes all support calls for software and hardware issues. For example, if you choose a Dell appliance, Dell takes all support calls and escalates to Nutanix for software support as needed. With a Nutanix NX appliance, all support calls go directly to Nutanix.

Software-only:

- Decouples software and support licensing from the underlying hardware, which enables:
  - License portability. You can use the same license even when the underlying hardware changes, as with a hardware vendor change or a node refresh.

    Note: Licenses are portable for like-for-like hardware replacements. If the hardware specification of the nodes changes, you might need to purchase additional licenses.

  - Deployment on additional supported hardware platforms, based on our qualified list of platforms, which you can find on the Hardware Compatibility Lists page.
  - Direct software support from Nutanix while the server vendor provides hardware support.
- Another type of software-only licensing is the core licensing option, which is best for customers who prefer the benefits of the software-only model but want hardware from a specific OEM server vendor. Core licensing enables customers to purchase software-only licenses and buy hardware from any of the appliance vendors. For example, XC-Core uses Dell XC OEM appliances but decouples software and hardware support.

Vendor Considerations

When it comes to selecting a hardware vendor, there are many factors to evaluate:

- Brand loyalty
- Support quality
- Hardware quality
- Operational experience
- Configuration options
- Physical form factor

Choosing Hardware

Most edge workloads can successfully run on any of the Nutanix and OEM models available, but some models and configurations may be a better fit than others. Once you know the requirements of your workloads, you can use Sizer (a Nutanix product that calculates node and cluster configuration based on inputs) to determine the best configuration for your clusters.

The following sections provide considerations for hardware model selection and performance.

Model Types

The various Nutanix and OEM appliance models and servers give you the flexibility to identify and use the right solution to meet financial, space, and performance requirements for different projects.

Servers for this solution fall under the edge category, which are similar to the servers used for general workloads and EUC but may offer fewer options for CPU and storage because they're optimized for edge-specific use cases.

Performance Considerations

Nutanix and AOS meet the performance demands of different workloads without continuous performance tuning. However, different cluster design and configuration options still yield performance benefits. Selecting the appropriate node model and configuration to meet application and solutions requirements is an important design decision. The primary design considerations for performance are:

- Number of drives: The number of HDDs or SSDs in a node can dramatically affect its performance characteristics.
  - Write-heavy workloads benefit from additional storage devices to provide performance and consistency.
  - Workloads like VDI typically have minimal capacity requirements but higher IOPS demands. It's common to use nodes with partially populated storage bays and as few as two flash devices per node, providing the right amount of storage capacity while still exceeding performance demands.

- All flash: All-flash clusters only use SSDs and provide higher IOPS and a more consistent I/O profile than HDDs.
- NVMe: New flash technology allows NVMe devices to deliver higher levels of I/O than SSDs with lower latencies.
  - To realize the full benefits of NVMe, nodes are typically configured with remote direct memory access (RDMA). RDMA allows one node to write directly to the memory of another node by allowing a VM running in the user space to directly access a NIC, which avoids TCP and kernel overhead resulting in CPU savings and performance gains.
- Size of flash tier: In hybrid configurations containing SSD and HDD devices, the bulk of the performance comes from the flash tier. The flash tier in hybrid clusters should be sized to meet or exceed the size of the working set for all applications running on the cluster. There's no penalty for having too much flash in a cluster but not having enough can result in inconsistent performance.

High-Level Design Recommendations

To summarize, for edge computing on Nutanix solutions, we recommend the following:

- Logically group the manageability, replication, and BCDR components of your edge sites in a single region or availability zone.
- Use Prism Central as the primary control plane.
  - If you use ESXi, you also need VMware vCenter.
- Use the simplified block-and-pod architecture detailed in the Block and Pod Architecture section of the Nutanix Hybrid Cloud Reference Architecture.
  - If you're building a large-scale site (more than four blocks with the expectation of additional blocks in the future) with the characteristics of a datacenter, see the Nutanix Hybrid Cloud Reference Architecture.

- Build clusters with one or three nodes.
  - If you need to be able to scale your edge solution beyond two nodes or add storage-only nodes to the cluster, you should implement a three-node cluster.
- If you use them, deploy storage-only nodes in pairs.
  - You can only use storage-only nodes in clusters with three or more nodes.
- Define an SOP that stipulates how to scale one- or two-node clusters in your design, as this step isn't a function of the core AOS and doesn't have a workflow.

5. Detailed Technical Design

This section provides technical guidelines for each layer of the design stack.

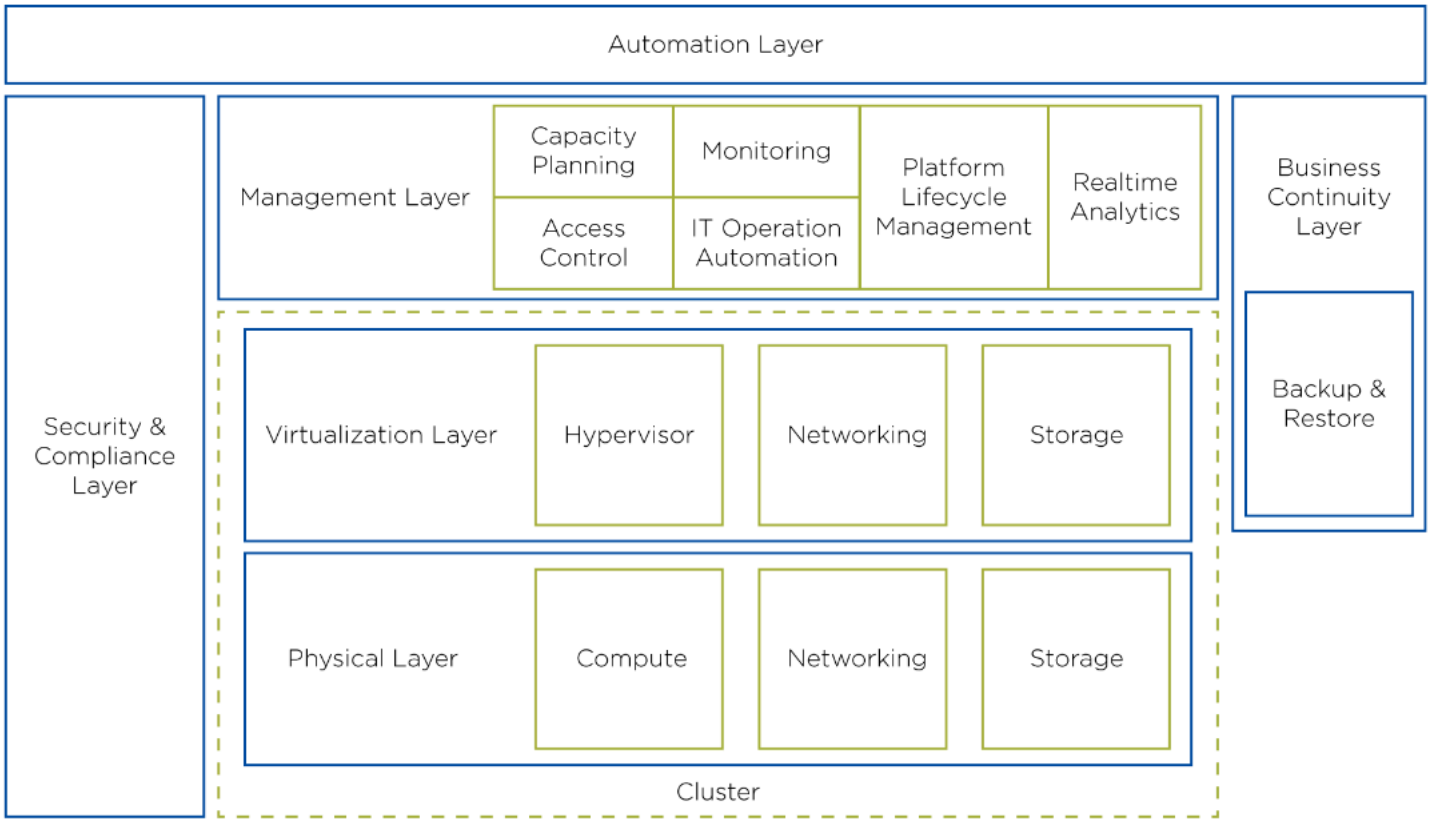

Figure 8: Overview of the Technical Design

Required Software Versions

We used the versions in the following table for this design.

Table: Software Versions

Name	Version	Reasoning
Prism Central	pc.2020.7	Use the latest version available.
Nutanix AOS	5.20	Use the latest LTS available.
Nutanix AHV	5.15	Use the latest LTS available.
VMware vSphere	6.7 or 7.0	Use either major release.
VMware vCenter	6.7 or 7.0	Recommended for vSphere deployments.

Physical Layer Design

This section guides you through the process of designing all physical aspects of your Nutanix edge deployment.

Choosing the Cluster Size

When designing a Nutanix cluster, it's important to consider hardware vendor recommendations, security, and operational requirements that may dictate the optimal cluster size. The following table provides a comprehensive list of the design considerations that pertain to cluster size.

Table: Cluster Size Design Considerations

Area	Limiting Factor	Considerations
Operations and manageability	Maintenance window	Define and validate your maintenance window and make sure the cluster upgrade process fits in the window. Example: If you have a maintenance window of 12 hours and a full single node upgrade takes at least 45 minutes, your maximum cluster size is 12–15 nodes.

Area	Limiting Factor	Considerations
Security and compliance	Security zones in an organization or business unit	Collect all relevant security and compliance requirements. Example: You have multiple security zones: internet perimeter, prod, test and dev, and inner perimeter. Every security zone has different cluster sizes: internet-facing perimeter clusters usually have fewer nodes and workloads to minimize impact of a security breach or DDoS attack. Less-critical test and dev clusters may also have more relaxed change management policies and can therefore have many hosts.
Vendor recommendations and limitations	Hypervisor limitations, management plane limitations, and vendor recommendations	Each product has limitations and vendor recommendations. Don't cross boundaries set by the vendor. Check the vendor's limitations and recommendations table.

Area	Limiting Factor	Considerations
Business continuity	RPO, RTO, and backup window	Collect BCDR requirements, RPO, RTO, backup and restore time window, and backup system performance statistics. Example: RPO: 24 hours, RTO: 48 hours. Ensure you can recover and restore from backup and restart workloads within 48 hours. A cluster where total storage capacity exceeds the technical capabilities of the backup system could fail to meet desired RTO.

Area	Limiting Factor	Considerations
Workload considerations	Application architecture, application licensing, and application criticality	Verify application architecture with the application team or vendor, including HA, disaster recovery, scale-in vs. scale-out, and performance requirements. For example: Application has its own HA or disaster recovery. If the application can provide native HA or disaster recovery, the RPO and RTO considerations described in the Business continuity row of this table may not apply. Consider the licensing model for each application and its implications. For example: Oracle or MS SQL licensing models are based on physical core count. Design clusters for database performance and capacity requirements to avoid cluster oversizing and minimize license costs.

Area	Limiting Factor	Considerations
Networking	Total available network switch ports and available network switch ports per rack	You need to know the available physical ports per rack and rack row when choosing cluster size and number of clusters. For example: A configuration with 96 × 10 Gbps ports available per rack, 48 × 1 Gbps ports available per rack, 2 × 10 Gbps uplinks per Nutanix host, and 1 × 1 Gbps uplink for out-of-band management can have at most 48 nodes per rack (total capacity of the top-of-rack switches).
Datacenter facility	Available server rooms, total power and cooling, power and cooling per rack, total available rack units, available rack units per rack, and floor weight capacity	Power and cooling are two of the most important factors limiting Nutanix cluster size and node density. Don't exceed any hard limits when designing your cluster layout. When calculating power consumption and thermal dissipation, use the maximum values provided by vendor. A typical datacenter rack is 42RU. Some datacenters have racks up to 58RU.
One- and two-node clusters	Planned vs. unplanned downtime and scalability	You need to know how to handle planned vs. unplanned events, and you need a plan for scaling beyond one node.

AHV deployments have the following limits:

- A scaled-out implementation of Prism Central on AHV can manage up to 400 clusters, 2,000 nodes, or 25,000 VMs (whichever limit is hit first).

  Note: One- and two-node AHV clusters are only available for ROBO deployments.

VMware vSphere deployments have the following limits:

- A scaled-out implementation of Prism Central on ESXi can manage up to 400 clusters, 2,000 nodes, or 25,000 VMs (whichever limit is hit first).
- VMware vCenter can support 2,000 hosts and 25,000 VMs (turned on).
- A single VMware ESXi cluster can support a minimum of 2 nodes and a maximum of 48.
  - The maximum number of nodes in a Nutanix cluster with ESXi is limited by the hypervisor version. For more details, refer to official VMware vSphere Documentation.

    Note: Two-node ESXi clusters are only available for ROBO deployments.

Failure Domain Considerations

Failure domains are the physical or logical parts of a computing environment or location that are adversely affected when a device or service experiences an issue or outage. The device or services affected greatly influence the size of the failure domain and its potential impact. For example, a router generally has a bigger failure domain than a wireless access point because more endpoints rely on a single router than a single access point. Identifying possible failure domains and keeping them small where possible reduces the chance of widespread disruption.

Building redundancy within and across failure domains helps mitigate the risks of failure. Nutanix clusters are resilient to a drive, node, block, and rack failures because they use redundancy factor 2 by default, allowing Nutanix clusters to self-heal.

Management Plane–Level Failure Domain

One of the most important failure domains is the management plane. The more workload domains a single management plane manages, the bigger the impact

of a failure. When deploying the management plane, consider the following risk mitigations to reduce the impact of a failure:

- At a minimum, design to meet the availability requirements of the managed workload or service with the highest uptime requirement.
- Confine workload domain to a single edge site and defined security zone.
- Ensure that the API gateway is always available because other third-party integrations (such as the backup vendor integration) may rely on it.
- Configure built-in RBAC to restrict access to management platform resources.
- Design your management plane to reside in the region or location where your edge clusters reside.

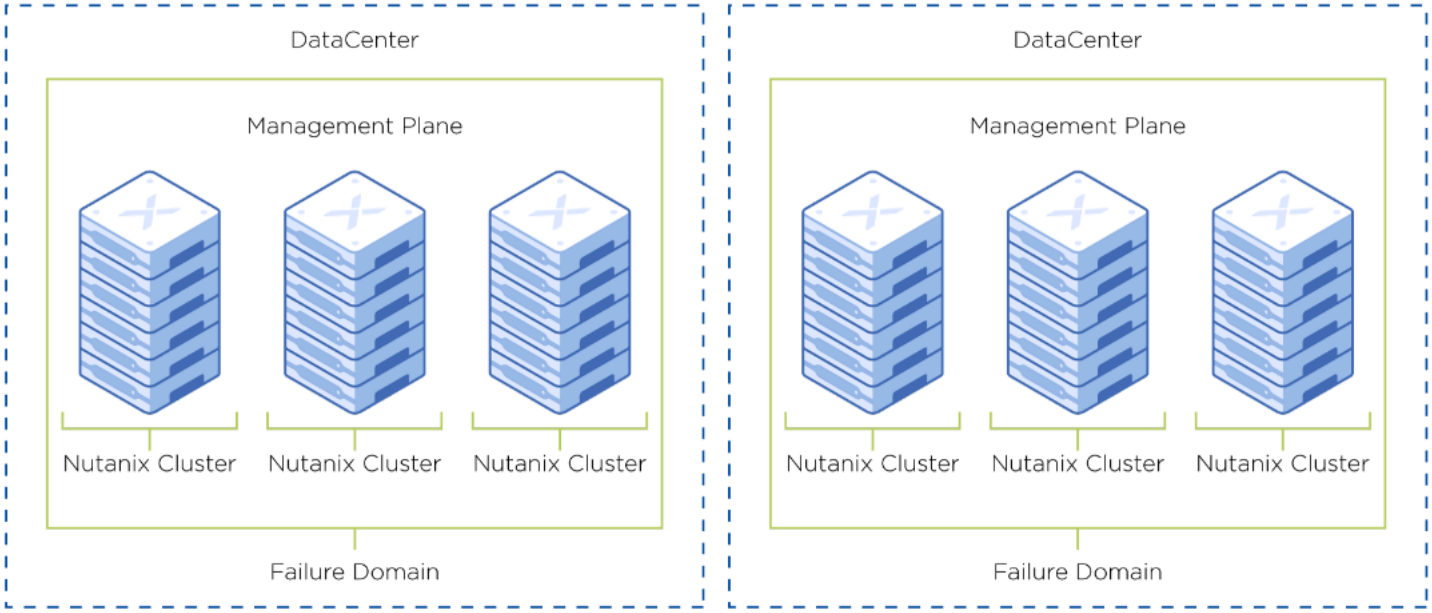

Figure 9: Management Plane Failure Domains

Cluster-Level Failure Domain

Nutanix clusters are another important failure domain. Large clusters result in larger failure domains and potentially higher business impacts, since they typically host considerably more workloads. To mitigate the risk of data

unavailability or service disruption, design for redundancy at the cluster level to protect data and services and apply the following mitigations:

- Use redundant power.
- Use redundant top-of-rack switches.
- Ensure redundant upstream connectivity to top-of-rack switches from each Nutanix node.
- Ensure redundant connectivity from edge to central datacenter.
- Use Nutanix scale-out architecture and data protection capabilities to replicate data to a second Nutanix cluster if one cluster fails.
- Deploy the application across multiple clusters.
- For smaller edge sites (fewer than five VMs) with cost constraints, consider implementing two single-node clusters because they operate as independent failure domains without the complexities of managing two-node clusters.

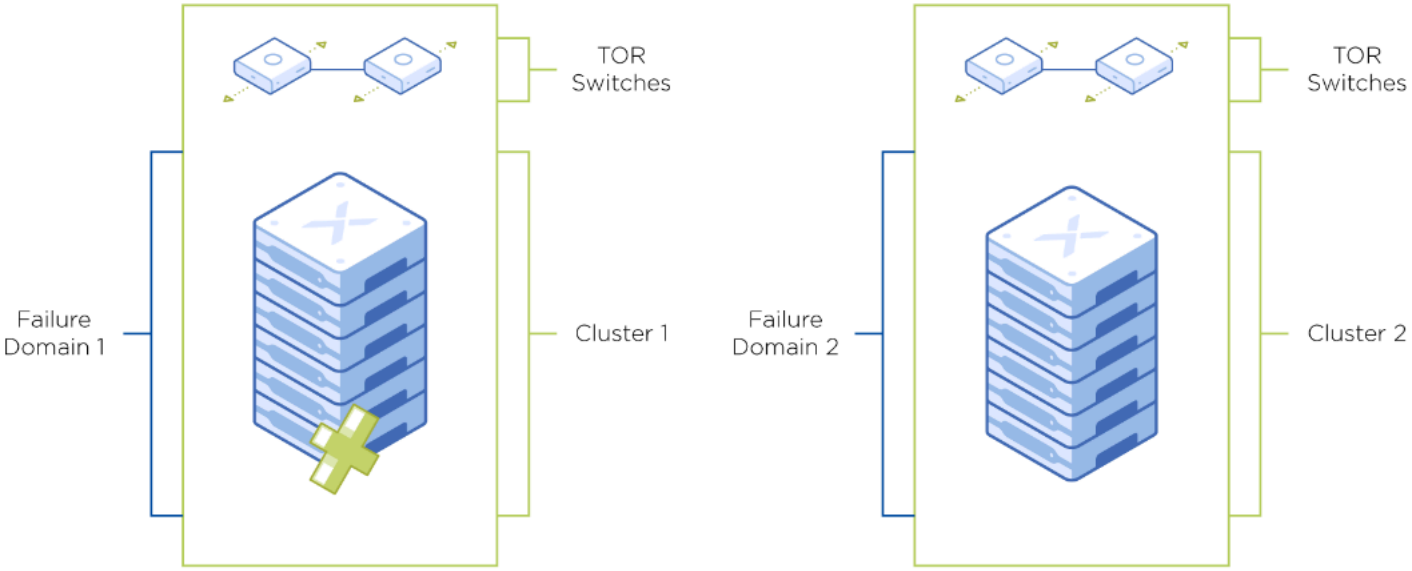

Figure 10: Nutanix Cluster Failure Domains

Rack and Server Room Failure Domains

Edge deployments can have a dedicated server room or a single wiring closet, but both represent a failure domain. Configure your deployment with the following mitigations:

- Use redundant power.
- Use independent cooling.
- Provide a redundant server room in a separate fire zone.
- Place the applications in multiple server rooms. For example, have AD domain controllers in separate server rooms.

If your edge site runs mission-critical workloads, consider mitigating risk by:

- Deploying stretched clusters across two server rooms in the edge site, representing a single failure domain with member nodes distributed across both rooms.
- Deploying two independent clusters across two server rooms in the edge site, representing two independent failure domains

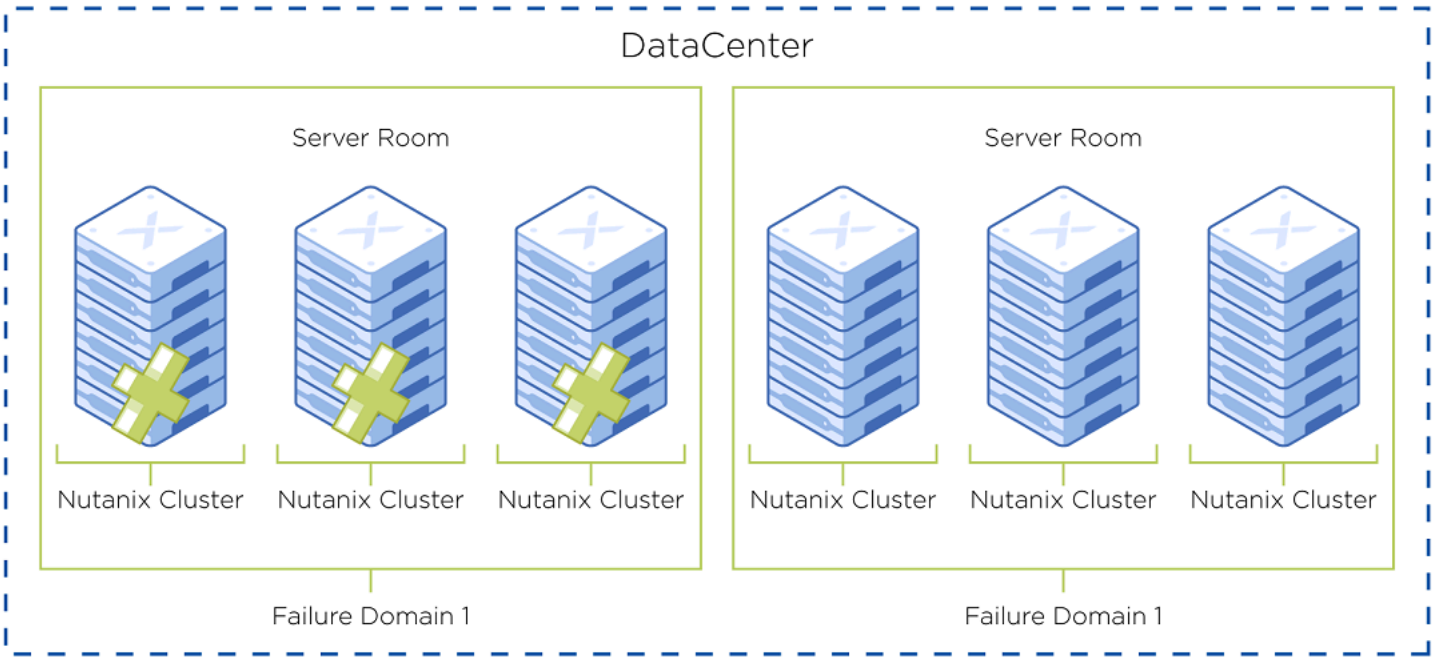

Figure 11: Server Room Failure Domains

Datacenter Failure Domains

Edge sites can lack the IT personnel necessary to assist during a failure scenario and keep the mean time to repair (MTTR) short. You should design your system with the following mitigations:

- Use redundant power.
- Use an independent and redundant cooling system for each datacenter room.
- Ensure redundant networking connectivity from different providers between datacenter buildings.
- Ensure redundant internet connectivity from different providers.
- Where possible, use multiple buildings and server rooms in separate buildings or campuses so that the failure of a single server room doesn't affect production. Distribute multicomponent services equally across datacenter server rooms.

- Ensure strict SLA compliance and keep well-documented resources and processes to help keep MTTR short.

Designing Workload Domains

This design uses a single workload domain at each edge site. Each workload domain consists of a set of Nutanix nodes managed by the same management instance (Prism Central or VMware vCenter) and connected to the same network domain.

A single workload domain can include different hardware and software combinations and have all ESXi or AHV nodes or mix ESXi (running VMs) and AHV (not running VMs) nodes in a single cluster. Single workload domains at edge sites provide the following advantages:

- Efficient space usage
- Default redundancy level
- Simple consumption
- Decreased capex and opex

> Note: Using a single workload domain increases the negative effect of a domain failure.

> Note: If your edge site runs multiple workloads with different requirements or constraints, consider using multiple workload domains. Refer to the Designing Workload Domains section in the Nutanix Hybrid Cloud Reference Architecture.

Networking for Edge Sites

A Nutanix cluster can tolerate multiple simultaneous failures and protect the cluster's read and write storage capabilities. However, this level of resilience requires a highly available, redundant network connecting a cluster's nodes. Even with intelligent data placement, if network connectivity between more than the allowed number of nodes breaks down, VMs on the cluster could experience write failures and enter read-only mode.

Consult the Physical Networking best practice guide to learn more about general networking guidelines and design requirements and recommendations.

Physical Switches

Edge sites operating as datacenters require datacenter switches with large buffers (10 Gbps or faster), which are critical in a large AOS cluster that sustains critical growth or hosts storage-intensive applications. In this case, see the Physical Switches section of the Nutanix Hybrid Cloud Reference Architecture.

In smaller edge clusters or ROBO deployments that have fewer than eight nodes or don't host write-intensive applications, the switch may not experience buffer contention, and you can relax these switch restrictions. The following switches don't meet high-performance datacenter switch requirements but are acceptable for these smaller edge clusters:

- Arista 7050
- Arista 7150S
- Cisco Catalyst 9000
- Cisco Catalyst 3000
- HPE FM2072

There are some switch types you should never use for any Nutanix deployment:

- Cisco Nexus 2000 (Fabric Extender)
- 10 Gbps expansion cards in a 1 Gbps access switch

Although Nutanix recommends an out-of-band management switch network separate from the primary network in traditional datacenter deployments, for enterprise edge sites these recommendations can be relaxed since the nodes typically have between one and three nodes. Configure IPMI, iLO, or iDRAC server-facing ports in the management network as access ports and don't use VLAN trunking for these ports. Restrict access to this critical management network.

Network Topology

In small edge sites running a handful of VMs, a simple network topology consisting of a pair of top-of-rack switches interconnected to a firewall and router is the simplest option. This simplified setup can address the requirements of a deployment with between one and three nodes.

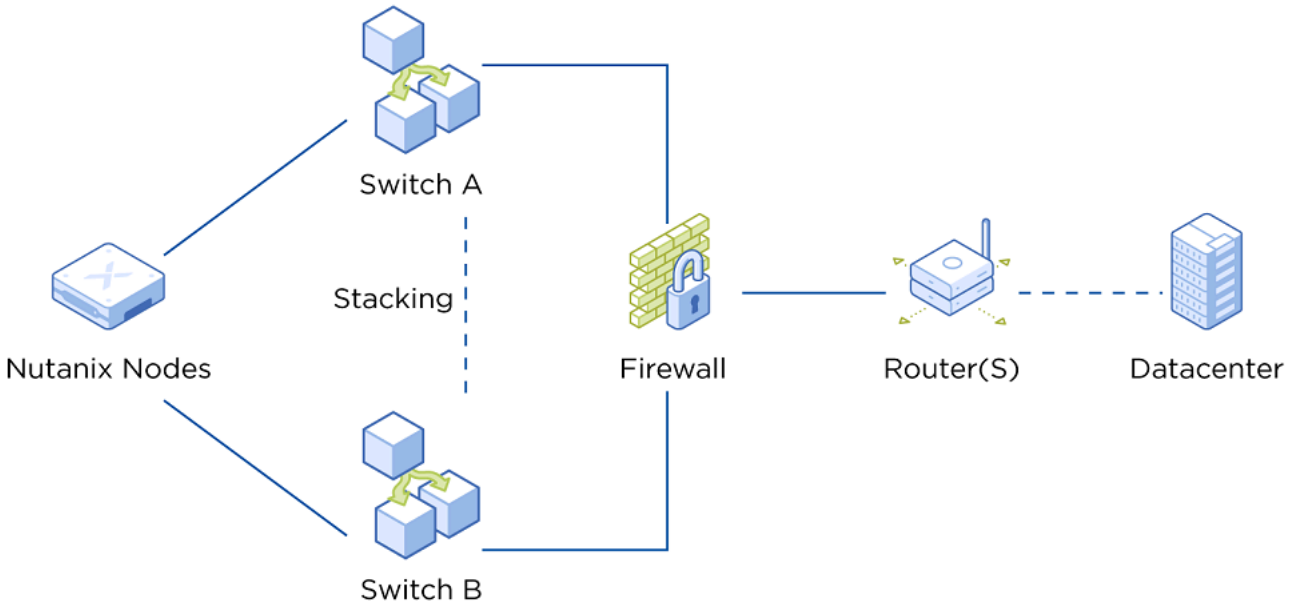

Figure 12: Small Edge Site Network Topology

For medium and larger deployments with multiple racks, Nutanix recommends using a leaf-spine topology because it's easy to scale, achieves high performance with low latency, and provides resilience. A leaf-spine topology requires at least two spine switches and two leaf switches. Every leaf connects to every spine using uplink ports.

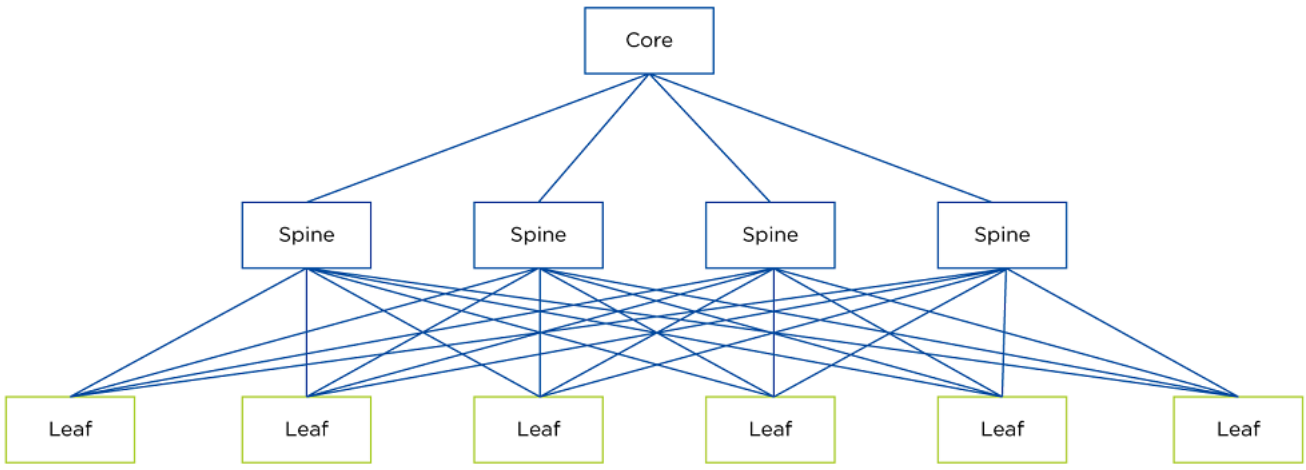

Figure 13: Leaf-Spine Topology for Medium and Larger Edge Sites

Note: If you're considering implementing a leaf-spine network topology, review the Network Topology section in the Nutanix Hybrid Cloud Reference Architecture for specific guidelines and recommendations.

The following list details Nutanix best practices and requirements for edge site network topology:

- Ensure that every Nutanix node in a cluster is in the same layer 2 broadcast domain and shares the same IP subnet.
- You can populate nodes designed exclusively for edge or ROBO with 1 Gbps networking for one- and two-node clusters with latency spikes of approximately 6 to 8 ms.
  - Limit your cluster size to at most eight nodes if you plan on using 1 Gbps.
- Ensure that when there are two or more nodes in a cluster, each layer of the network topology is highly available and tolerates individual device failures.
  - Avoid configurations or technologies that don't maintain system availability during single-device outages or upgrades, such as stacked switches.
- Ensure that there are no more than three switches between any two Nutanix nodes in the same cluster.
- Don't use WAN or remote links between Nutanix nodes in the same Nutanix cluster.
- Separate Nutanix CVM and hypervisor hosts in a dedicated VLAN that doesn't include any VM traffic.
- Don't place Nutanix nodes in the same Nutanix cluster if the stretched layer 2 network spans multiple edge sites, buildings, datacenters, or availability zones or if there's a remote link between the two locations.
  - Using a stretched layer 2 network over a layer 3 network is only acceptable when the Nutanix cluster is in the same switch fabric or aggregation layer, such as when a layer 2 network stretches between two racks in the same datacenter.
- Don't use features like block or rack awareness to stretch a Nutanix cluster between different physical sites.

- Configure adequate uplinks between switches or interswitch links for east-west storage traffic to minimize port-to-port oversubscription. For example, use multiple 40 Gbps uplinks (or interswitch links).
- Connect hosts using redundant links.
- Configure switch ports facing Nutanix servers as spanning tree portfast or edge to skip the listening and learning phases and prevent cluster outages caused by changes to spanning tree topology.
- Configure the CVM and hypervisor VLAN as native, or untagged, on server-facing switch ports. Newly added nodes use untagged traffic for discovery and work out of the box, reducing manual server configuration.
- Use tagged VLANs on the switch ports for all guest workloads to keep the workloads separate from each other and from the CVM and hypervisor network.
- Reduce network oversubscription to ensure as close to a one-to-one ratio as possible. Dropped network packets or a congested network immediately affect storage performance.

Broadcast Domains

Nutanix recommends a traditional layer 2 network design to ensure that CVMs and hosts can communicate in the same broadcast domain even if they're in separate racks. The CVM and host must be in the same broadcast domain and IP subnet.

> Note: If you choose a layer 3 network design, refer to the Broadcast Domains section of the Nutanix Hybrid Cloud Reference Architecture.

Network Scalability

Because edge sites are typically small, they contain a simplified core, aggregation, and access layer switch design. In this scenario, to scale, you simply add additional access layer switches, taking into consideration the oversubscription ratios in the design.

If you used a leaf-spine network topology, see the Scaling the Network section of the Nutanix Hybrid Cloud Reference Architecture.

Connectivity to the Core Datacenter

There are multiple ways to establish connectivity from the edge to the core datacenter, including traditional VPNs and direct connect methods. Evaluating the various advantages and disadvantages of each method is beyond the scope of this document; however, keep in mind the latency and bandwidth requirements for edge sites:

- Witness VMs hosted in the cloud can sustain up to 500 ms of latency per 100 instances they manage.
- Prism Central requires at least1.5 Mbps and a maximum latency of 150 ms to run CRUD operations, collect statistics, and perform updates.
- Scheduled replications have no set bandwidth requirements.

AHV Networking

AHV uses Open vSwitch (OVS) for all VM networking. The virtual switch is referred to as a bridge and uses a br prefix in the name. AHV also includes a Linux bridge called virbr0. The virbr0 Linux bridge carries management traffic between the CVM and AHV host. All other storage, host, and workload network traffic flows through the br0 OVS bridge or additional brN bridges if configured.

Read the AHV Networking best practice guide for in-depth guidance on any settings not covered here.

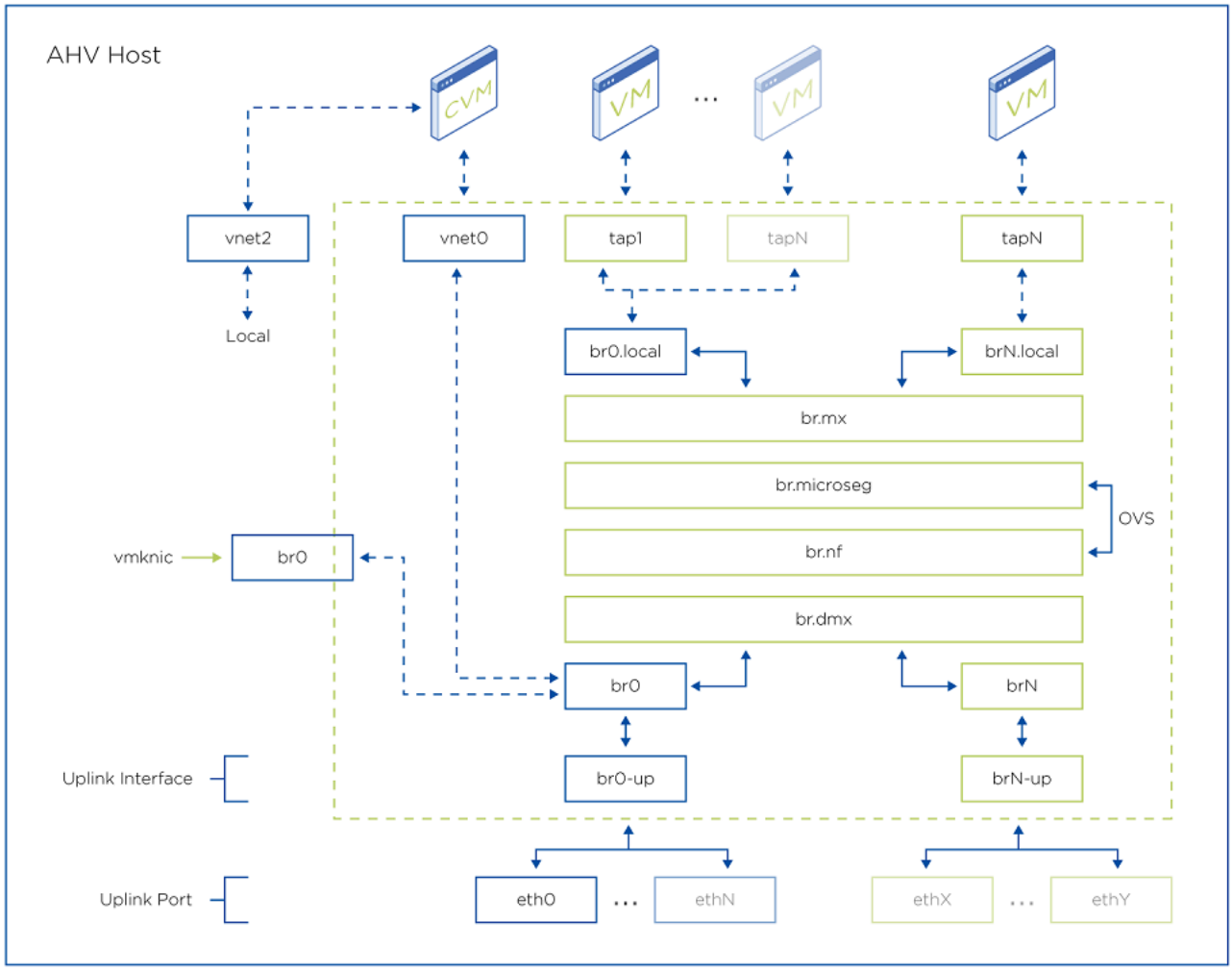

Figure 14: AHV Networking

The following list details the Nutanix best practices and requirements for edge site AHV networking:

- Use only the default bridge, br0, with at least two of the fastest network uplink adapters of the same speed to simplify the design.
  - Converge the management, storage, and workload traffic on this single pair of uplink adapters.

- Only add additional brN bridges when you need a connection to a separate physical network. For example, if the top-of-rack switch has two pairs of switches—one pair for storage and management and another pair for workload traffic—it makes sense to create another bridge, br1, and place the workloads on this bridge.
- Don't modify the configuration of any bridges inside the AHV host unless following an official Nutanix guide.
- To keep network configuration simple and reduce risk, use the standard 1,500-byte MTU in the hosts, CVMs, and workload VMs. Nutanix doesn't recommend jumbo frames unless specifically required by high-performance Nutanix Volumes iSCSI workloads or specific workload requirements.
  - When switching from 1,500-byte frames to 9,000-byte frames, performance improvements are generally not significant unless the workload uses the maximum network bandwidth for read traffic. For more information on when to use jumbo frames, see the Nutanix Volumes best practice guide and AHV Networking best practice guide.
- Connect at least one 10 Gbps or faster NIC to each tor switch to maintain high availability if one switch is lost.
- Use NICs from the same vendor within a bond to ensure compatibility and prevent undesired failover behavior.
- Use VLANs to separate logical networks. Physical hosts have a limited number of network ports and each port adds complexity. You can separate traffic logically without numerous physical ports.
- To simplify the design, use active-backup uplink load balancing.

vSphere Networking

VMware vSphere networking follows many of the same design decisions as AHV networking. Nutanix hosts with vSphere ESXi use two virtual switches (vSwitches), named vSwitchNutanix and vSwitch0. vSwitchNutanix is the internal, standard vSwitch used for management and storage traffic between the CVM and the hypervisor. vSwitch0 is also a standard vSwitch by default, used for communication between CVMs and workload traffic.

The critical design choices for vSphere networking are covered here, and you can refer to the VMware vSphere Networking best practice guide for more details.

The following list details the Nutanix best practices and requirements for edge site vSphere networking:

- Don't modify vSwitchNutanix.
- Convert vSwitch0 to the virtual distributed switch (VDS) following the instructions in Migrate from a Standard Switch to a Distributed Switch. Converting to the VDS allows you to centrally manage networking for all hosts, instead of configuring each host network individually. The VDS also enables advanced networking functions such as load-based teaming, LACP, and traffic shaping.
- To simplify the design, connect at least two of the fastest adapters of the same speed to vSwitch0 and use the Route Based on Physical NIC Load load balancing method to ensure that traffic is balanced between uplink adapters.
  - Connect these adapters to two separate top-of-rack switches to ensure redundancy.
- Connect at least one 10 Gbps or faster NIC to each top-of-rack switch to maintain high availability if one switch is lost.
- Don't add more vSwitches unless you need to connect to another physical network to meet security or workload requirements.
- All CVM storage, hypervisor host, and workload traffic should flow through vSwitch0, using VLANs to separate the workload traffic and all other traffic.
- Use the default 1,500-byte frame size on all uplinks unless there is a specific performance or application requirement that justifies 9,000-byte jumbo frames.

Virtualization Layer Design

This section describes technical design considerations for Nutanix AHV and VMware ESXi.

Nutanix AHV

AHV delivers virtualization capabilities for the most demanding workloads, and provides an open platform for server virtualization, network virtualization, security, and application mobility. When combined with comprehensive operational insights and virtualization management from Prism and Prism Central, AHV provides a complete datacenter solution.

Nutanix AHV has built-in VM high availability (VMHA) and a resource contention avoidance engine called Acropolis Dynamic Scheduling (ADS). ADS is always on and doesn't require any manual tuning. There are two main levels of VMHA for AHV: best effort (no node or memory reservations required) and guarantee (some memory reserved on each node to enable failover). Use the guarantee level for VMHA in production environments and best effort in test and development environments. For additional information, see the Virtual Machine High Availability tech note.

Like VMware's Enhanced vMotion Capability (EVC), which allows VMs to move between different processor generations, AHV determines the lowest processor generation in the cluster and constrains all QEMU domains to that level. This feature enables you to mix processor generations in an AHV cluster and to live-migrate between hosts.

VMware vSphere

If you use vSphere in your enterprise edge deployment, deploy a highly available vCenter instance with an embedded Platform Services Controller to manage the ESXi-based Nutanix clusters. Nutanix upgrade automation features, such as one-click upgrades and LCM, require the advanced control features that vCenter provides. You also need vCenter Server to create and manage the VMware vSphere cluster responsible for the Distributed Resource Scheduler (DRS) and HA, and Prism Central to manage the Nutanix components.

> Note: If you're using the disaster recovery runbook automation feature, you need to deploy separate vCenter and Prism Central instances in each region or availability zone.

EVC allows VMs to move between different processor generations in a cluster. EVC mode is a manually configured option, unlike the corresponding feature in AHV. Enable EVC at cluster installation to help ensure that future node additions

are seamless. Set EVC mode to the highest compatibility level the cluster processors support.

VMware HA and DRS are core features you should use in your edge site design. Nutanix best practices specify a few HA and DRS configuration changes from the default:

- Enable HA to automatically restart VMs in case of node failure.
- Enable DRS with the default automation level so that it can move VMs as needed to ensure optimal performance.

  Note: Moving VMs may temporarily affect Nutanix data locality.

- Configure DAS.IGNOREINSUFFICIENTHBDATASTORE if one Nutanix container is presented to the ESXi hosts to eliminate false positives when a cluster uses a single datastore.
- Disable automation level, HA, and DRS for all CVMs. HA doesn't need to reserve resources for CVMs because they're bound to a single node and can't be restarted elsewhere in the cluster.
- Set Host failures cluster tolerates to 1 for replication factor 2 and 2 for replication factor 3. These settings ensure that the proper resources are automatically reserved for the cluster.
- Set the host isolation response to Power off and restart VMs to ensure that VMs are moved to a healthy host and continue to function.
- Set the host isolation response to Leave powered on for CVMs. You don't want CVMs to be turned off if there's a transient network disruption.
- Disable Storage I/O Control. If Storage I/O Control is enabled, it can cause storage unavailability, unnecessary lock files, and complications with Metro Availability.

Management Layer Design

Prism Central

There are two deployment architectures for Prism Central that you can use and scale depending on the size and goals of the design:

1. Single-VM Prism Central: One instance in each region or availability. Required if you use the disaster recovery runbook automation feature, which requires that the source and target sites have separate Prism Central instances controlling them. Requires additional cluster resources to run the required VMs.
2. Scaled-out Prism Central cluster: Three-VM cluster. Increases management plane availability and the total number of objects you can manage. Requires additional cluster resources to run the required VMs.

We recommend deploying the scaled-out Prism Central instance for edge sites to simplify capacity planning, platform lifecycle management, and virtual networking management, and to reduce management overhead.

VMware vCenter Server

You can deploy vCenter in many different sizes, and the size ultimately determines how large an environment you can support in terms of VMs, hosts, and clusters. Refer to the official VMware documentation for the version you're deploying to determine the proper size for your design.

You can deploy vCenter as a single VM or as a vCenter High Availability (vCenter HA) instance. vCenter HA doesn't increase the size of the environment it can manage because a single virtual appliance is active at any time. The size of the environment that can be managed is based on the size of the VMs deployed.

We recommend deploying one instance of vCenter HA for all workloads and configuring it with Platform Services Controller embedded.

Dependent Infrastructure

There are a variety of other infrastructure services that are necessary for a successful Nutanix deployment, such as NTP, DNS, and AD. You may already

have these infrastructure services deployed and available for use when you deploy your Nutanix environment, but if you don't, you need to deploy them as part of the new environment.

NTP

If clock times drift too far apart, some products may have trouble communicating across layers of the solution. Network Time Protocol (NTP) synchronizes computer clock times including network, storage, compute, and software.

The following list details the Nutanix best practices and requirements for NTP for edge sites:

- Configure at least three NTP servers (NTP standard recommendation is five to detect rogue time sources) and ensure that they're accessible at all solution layers, including AOS, AHV, and Prism Central, plus vCenter and ESXi if you're using vSphere.
- Use the same NTP servers for all infrastructure components.
- Don't use an AD domain controller as an NTP source.
- If you're in a dark site with no internet connectivity, use a switch or GPS time source.
- Configure NTP sources specific to the region where the edge clusters reside.

DNS

Domain Name System (DNS) is a directory service that translates domain names of the form domainname.ext to IP addresses. DNS ensures that all layers can resolve names and communicate.

We recommend configuring at least two DNS servers and making them accessible at all layers (AOS, Prism Central, ESXi, vCenter, and network switches) to ensure that components can reliably resolve addresses at all times. We also recommend configuring DNS sources specific to the region where the edge clusters reside.

Active Directory

Active Directory (AD) often serves as the authoritative directory for all applications and infrastructure in an organization. For this design, all the consoles and element managers use RBAC and use AD as the directory service for user and group accounts. Where possible, use AD groups to assign privileges for easier operations. You can then control user access by adding or removing a user from the appropriate group.

Logging Infrastructure

Capturing logs at all layers of the infrastructure is very important. If there is a security incident, logs can be critical for forensics. An example of a robust log collector is Splunk, but there are other options.

The following list details the Nutanix best practices and requirements for logging infrastructure in edge sites:

- Deploy a robust logging platform that meets your security requirements.
- Forward all infrastructure logs to the centralized log repository.
- Store the logs in a different cluster, or location, from where you collect them. This measure protects the logs in case of a catastrophic cluster failure, ensuring they can later be used for forensics.
- Ensure that AD sites and services are set up to define a replication and authentication topology if connectivity to the datacenter is lost.

Security Layer Design

You can build private or multitenant solutions on the Nutanix Cloud Platform, so the security responsibilities vary based on the use case. The security approach includes multiple components:

- Physical
- Virtual infrastructure
- Threat vectors
- Workloads

We designed Nutanix infrastructure to deliver a high level of security with less effort. Nutanix publishes custom security baseline documents for compliance, based on United States Department of Defense (DoD) RHEL 7 Security Technical Implementation Guides (STIGs) that cover the entire infrastructure stack and prescribe steps to secure deployments in the field. The STIGs use machine-readable code to automate compliance against rigorous common standards.

Nutanix provides SCMA by default. SCMA checks multiple security entities for both Nutanix storage and AHV, and Nutanix automatically reports inconsistencies and reverts them to the baseline. With SCMA, you can schedule the STIG to run hourly, daily, weekly, or monthly. The STIG has the lowest system priority in the virtual storage controller, ensuring that security checks don't interfere with platform performance.

In addition, Nutanix releases Security Advisories that describe potential security threats and their associated risks plus any identified mitigations or patches. For more detailed information on Nutanix security and compliance features, see the following documents, separated by category:

- Networking: Flow Network Security tech note
- Information: Information Security tech note
- Cloud: Nutanix DRaaS Security tech note
- Database: Nutanix Era Security tech note

Authentication Best Practices

- Maintain as few user and group management systems as possible. A centrally managed authentication point is preferred to many separately managed systems.
- You should at least take advantage of the external LDAP support provided by Nutanix components.

- Use AD authentication for user and server accounts to ensure that user activity logged for auditing purposes and account security is configured and maintained from a single centralized solution.
  - › You must have a highly available AD infrastructure and network connection to the datacenter.
- Use an SSL or TLS connection to AD to eliminate cleartext exchanges on the network.

Certificate Best Practices

- Protect all consumer-facing components with certificates signed by a trusted certificate authority to provide an extra layer of security and prevent meddler-in-the-middle (MITM) attacks.
  - › You can use internally or externally signed certificates based on your consumer classification and the service the specific component provides.

Cluster Lockdown Best Practices

Don't use Nutanix or vSphere cluster lockdown (feature that lets you enforce SSH access to CVMs and host using key pairs instead of passwords) unless you require passwordless communication. This decision helps restrict direct access to the CVM and hypervisor to as few entities as possible.

Hardening Best Practices

- Enable Advanced Intrusion Detection Environment (AIDE) for the CVM and AHV. AIDE performs checksum verification for all statis binaries and libraries for improved security.
- Enable stack traces for cluster issues for AHV or ESXi and the CVM.
- Enforce a complex password policy (at least 15 characters long with at least 8 different characters) for the hypervisor and CVM.
- Enable a banner for AHV or ESXi and the CVM that retrieves a specific sign-in message via SSH.
- Enforce only SNMPv3 on the CVM.
- Configure SCMA to run hourly to capture unacceptable configuration drift.

- ESXi only: Stop unused services and close unused firewall ports to limit the attack surface.

  Note: Ensure that you don't stop a service or close a firewall port that's required by Nutanix, such as SSH and NFS.

- Generally, we recommend following the Hardening Controller VM section of the AOS Security Guide. If you're using AHV, follow the Hardening AHV section of the AOS Security Guide.

Internet-Facing Services Best Practices

We recommend using multiple internet connections (active-backup) and ISP-provided denial-of-service (DoS) and distributed-denial-of-service (DDoS) filtering to help mitigate the potential effects of a DoS or DDoS attack.

Logging Best Practices

- Configure and maintain a single centralized logging solution for auditing purposes and account security.
- Send log files to a highly available syslog infrastructure.
- At least one of the individual targets of the highly available logging infrastructure should run outside the virtual infrastructure itself so that if the virtual infrastructure is compromised, forensic investigations can access logs that might not be available on the cluster itself.
- Include data from all Nutanix modules and components in logging using the error-log level and ensure that they're searchable. See the Configuring the Remote Syslog Server Settings section of the Acropolis Advanced Administration Guide for more information.
- Use default ESXi logging levels, log rotation, and log file sizes.
- If you have additional security and reliability requirements, use TCP for log transport. Otherwise, use the default syslog protocol, UDP.
- Use port 514 (defined port in syslog RFC) for logging.

Network Segmentation Best Practices

- To protect Nutanix CVM and hypervisor traffic, place them together in their own dedicated VLAN, separate from other VM traffic. This configuration applies to all hosts in a cluster or single node.
- Place out-of-band management on a separate VLAN or physical network to provide additional security.
- We recommend configuring the CVM and hypervisor host VLAN as a native, or untagged, VLAN on the connected switch ports to allow easy node addition and cluster expansion.
- Don't segment Nutanix storage and replication traffic, or iSCSI Volumes traffic, on separate interfaces (VLAN or physical) unless additional segmentation is required by mandatory security policy or the use of separate physical networks.

Role-Based Access Control (RBAC) Best Practices

- Use a least-privilege and separation-of-duties approach when assigning RBAC permissions to ensure that each group or individual user has just enough permissions to perform their duties. Use predefined roles or create new roles as needed.
- Configure RBAC at the Prism Central level because it provides the overlying management construct. This configuration ensures that your least-privilege configuration stays in place and avoids common mistakes that occur when RBAC is configured at multiple different levels.
- Align RBAC structure and default plus custom roles with your company requirements.

Data-at-Rest Encryption Best Practices

> Note: AHV encrypts the entire Nutanix cluster, while ESXi gives you the option to define encryption on a per-Nutanix-container basis if required.

- Keep management traffic, including storage traffic, on a separate network.
- Don't use storage encryption.

- Don't use a key management server (KMS).

> Note: All methods of DaRE are FIPS 140-2 compliant. However, if you require levels 2, 3, or 4, you also need a hardware component.

Microsegmentation and Firewall Best Practices

- Use Flow Network Security (previously Flow Microsegmentation) to comply with regulatory and business-specific compliance policies such as PCI-DSS, HIPAA, and NIST.
- Use a Palo Alto Networks VM-Series firewall with Flow Network Security to provide advanced threat and vulnerability detection for layers 4–7.

For more information on Flow Network Security, see the Flow Network Security software documentation and the Flow Network Security tech note. For additional information on deploying the Palo Alto VM-Series on AHV, see the Palo Alto Networks VM-Series Firewalls on Xi tech note and the Nutanix and Palo Alto Networks Service Chain Integration Guide.

Automation Layer Design

Nutanix supports intelligent IT operations and advanced automation that enable you to streamline operations, enable IT as a service, and support the needs of developers and business teams. This section covers virtual infrastructure automation and orchestration, focusing on provisioning and maintenance.

Life Cycle Manager (LCM)

Nutanix LCM determines any software and firmware dependencies, intelligently prioritizes updates, and automates the entire upgrade process across all clustered hosts, without any impact to applications or data availability. It supports one-click upgrades across multiple qualified hardware manufacturers and configurations, so IT teams have the flexibility to deploy the best hardware for each use case and still benefit from centralized upgrade capabilities.

LCM offers the best software and firmware upgrade experience for clusters with three or more nodes. In this design, the nodes can locally coordinate data synchronization, VM evacuation and migration, and node recovery and

availability during the upgrade process for the firmware and software on the individual nodes. The individual nodes might restart multiple times, and that entire process is automated through LCM without affecting applications and data.

Certain LCM features aren't supported when your cluster contains one or two nodes:

- In a one-node scenario, the LCM software and firmware upgrades aren't supported because there's no hardware redundancy. In this design, the user must plan their downtime and trigger the software upgrade manually via Prism or use the respective hardware manufacturer's management tools.
- In a two-node scenario, LCM software and firmware is supported, but during the upgrade process, data synchronization across nodes takes more time because there's only one other node in the cluster to absorb the changes. You must complete witness upgrades separately from data-carrying node upgrades. For this reason, we recommend a larger maintenance window for two-node clusters.

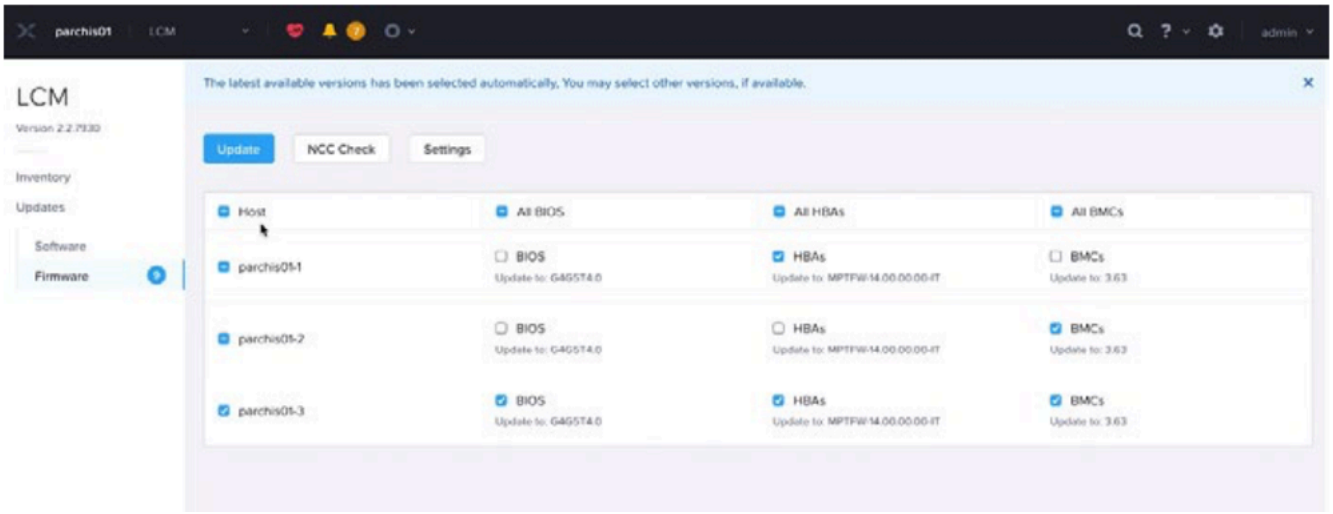

Figure 15: Life Cycle Manager

Foundation

Foundation manages the initial setup and configuration of a cluster. Nutanix nodes can come with AHV and the CVM preinstalled, and with Foundation you can:

- Add the nodes to an existing Nutanix cluster.
- Create a new Nutanix cluster.
- Reimage the nodes with different a AHV and or AOS version or different hypervisor and create a Nutanix cluster.

> Note: The Foundation process may vary for different hardware vendors.

Nutanix provides a Foundation preconfiguration service that you can install from nutanix.com. With the service, you can define and download the Nutanix cluster configuration to be used during the Foundation process. The downloaded file contains all configuration required to perform the Foundation operation and comes in json format, which makes it easy to keep track of configurations, document them, and share with your peers.

See the Field Installation Guide for the functions available for the different Foundation software options.

Foundation Central

Foundation Central enables you to create clusters from factory-imaged nodes and reimage existing nodes that are already registered with Foundation Central remotely from Prism Central. This feature enables you to create clusters on remote sites without having to arrange a personnel visit.

Foundation Central has the following limitations:

- Foundation Central supports creation of a one-node cluster without imaging on factory-shipped nodes with AOS and any hypervisor installed.
- Foundation Central supports creation of a one-node cluster with imaging only on nodes that are shipped with DiscoveryOS installed.
- Imaging logs for the remote nodes aren't available on Prism Central. You can only access these logs in the nodes.

- Foundation Central doesn't support updating the CVM memory and hostname for one-node cluster creation without imaging.

Nutanix Calm

Calm provides infrastructure and application automation and life cycle management for the Nutanix enterprise cloud and public clouds. It provisions, scales, and manages infrastructure and applications across multiple environments to make the entire IT infrastructure more agile and application-centric. Building an enterprise or service provider cloud solution with Nutanix Calm helps organizations with delivery and ongoing management of infrastructure, applications, and custom services, leading to lower opex.

We recommend reviewing the Nutanix Calm Reference Architecture guide for further information on implementing Nutanix Calm

Operations Design

Note: We recommend purchasing the Nutanix Cloud Manager (NCM) Starter license, which gives you access to analytics, capacity planning, custom dashboards, and playbooks for Prism Central.

Native runway calculations built into Prism Central automatically calculate the remaining capacity of the system as soon as the Prism Element instance is brought under Prism Central management. Configure these runway calculations to run as part of periodic reports and review them on a regular basis (at least monthly) to ensure that sufficient capacity exists. This configuration is especially important for organizations that have a significant lag between when they commit to purchase additional gear and when the gear is online and available to use.

Figure 16: Capacity Runway

Accurate and efficient system-level planning requires accurate sizing for individual workloads. Machine learning in Prism Central provides anomaly detection for VMs when the workload crosses learned thresholds. The system also categorizes VMs based on their behavior, and you can create custom alert policies that match VMs by conditions.

Operations Best Practices and Requirements

- Review capacity planning monthly.
- Perform updates after hours for performance- or migration-sensitive applications to reduce the load on the system.
- Use the current Long-Term Support (LTS) branch unless you require a specific new feature for your design.
- Update to the next maintenance version four weeks after release and to the current patch version two weeks after release to keep updates relatively small.
- Maintain a preproduction environment to test any necessary changes to firmware, software, or hardware before implementing the change in production. Preproduction environments can help you avoid the cost of outages.

- Configure alerts and alert policies in Prism Central, not Prism Element, to create consistency across multiple clusters and reduce the effort involved in making multiple changes. Prism Central–based alert policies also allow you to detect anomalies.
- Use SMTP for alert transmission to create consistency across multiple clusters, reduce the effort involved in making multiple changes, enable anomaly detection, and offer more options for delivery.

BCDR Design

Nutanix provides a range of data protection features. Our general recommendation for BCDR design is: If an application has native data protection mechanisms (Exchange database availability groups (DAG), SQL Server Always On, Oracle RAC), use those to provide application BCDR. If an application doesn't provide a data protection mechanism, use the built-in Nutanix protection mechanisms.

Consult the Multi-Datacenter Design for Nutanix Core and BC/DR section of the Nutanix Hybrid Cloud Reference Architecture for considerations and strategies for multi-datacenter designs.

BCDR Logical Implementations for Edge Sites

Disaster recovery patterns for edge sites vary from organization to organization, but one-to-one and many-to-one implementations are the most common.

One-to-One

For smaller environments with only a pair of locations, protect applications between two availability zones in the same region or across regions. You can protect the same workload or application with different RPO thresholds. The availability zone location you choose (same region or across regions) defines the RPO value for the protected application or workload: If an application or workload requires RPO 0, choose availability zones in the same region (latency requirement for RPO 0 is 5 ms RTT or less).

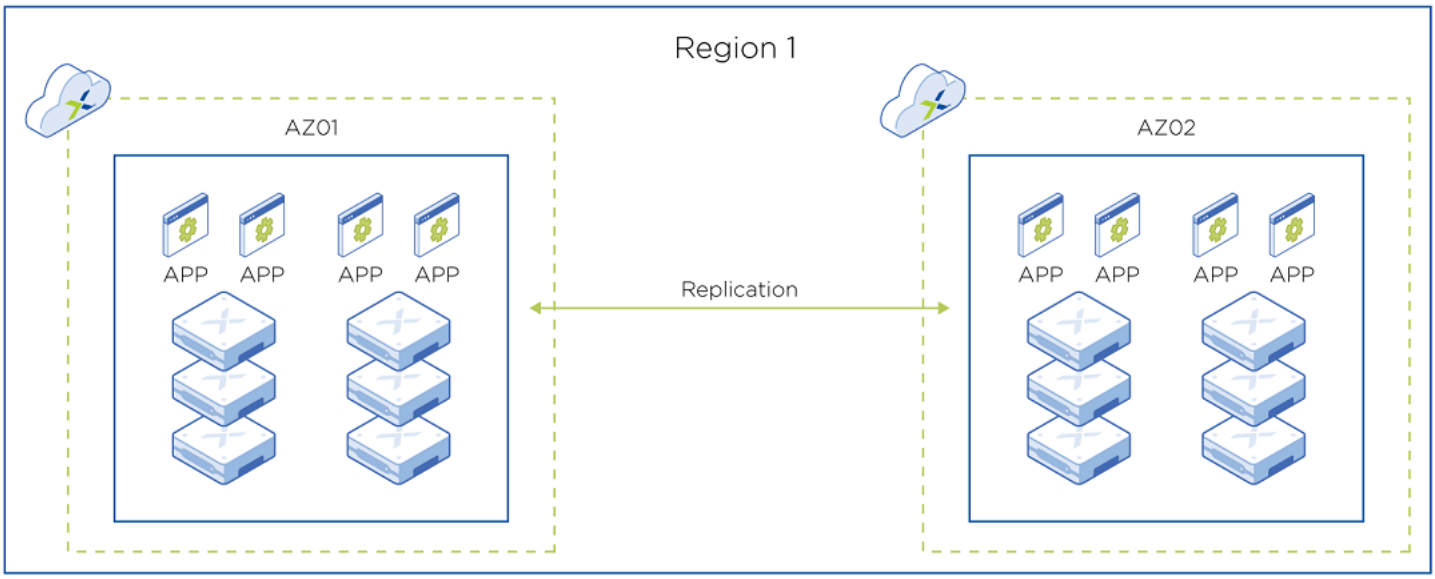

Figure 17: Two Availability Zones in the Same Region

Many-to-One

The many-to-one pattern maps multiple source edge clusters to single disaster recovery target Nutanix cluster or core datacenter. In the most common configuration, each edge location is a separate availability zone that replicates data to the disaster recovery hub, which is located in a separate availability zone in the same region.

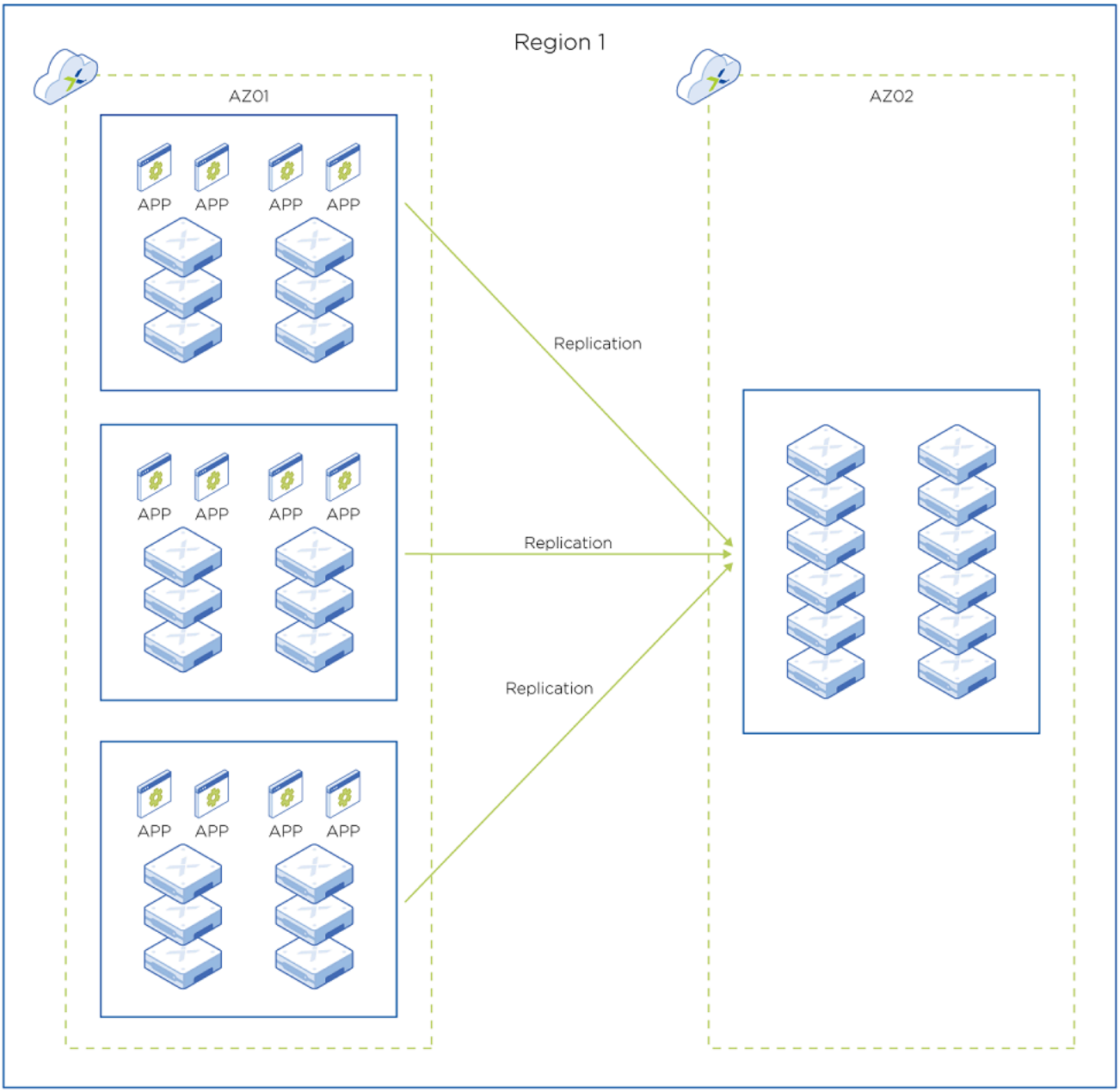

Figure 18: Multiple Availability Zones in a Single Region

You can also have multiple availability zones across multiple regions.

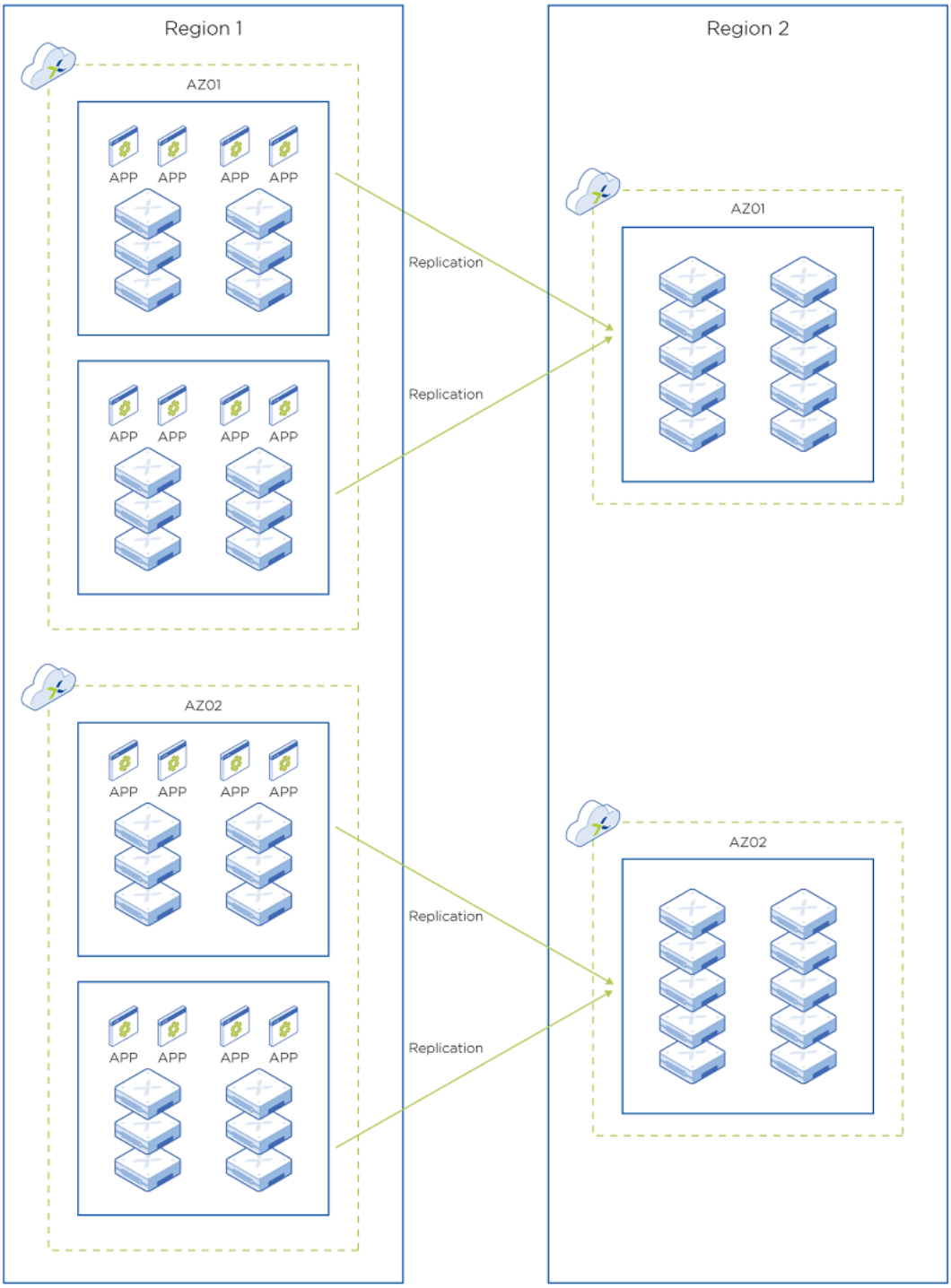

Figure 19: Multiple Availability Zones across Multiple Regions

Note: Consult the Nutanix Hybrid Cloud Reference Architecture to evaluate other implementations.

Nutanix BCDR Solutions

A good disaster recovery solution involves different levels of service because applications have different levels of criticality to the business and therefore different disaster recovery requirements. For this reason, this document includes multiple architectures for a variety of scenarios. The following flow chart walks through the decision tree for choosing the best Nutanix disaster recovery solution for your requirements.

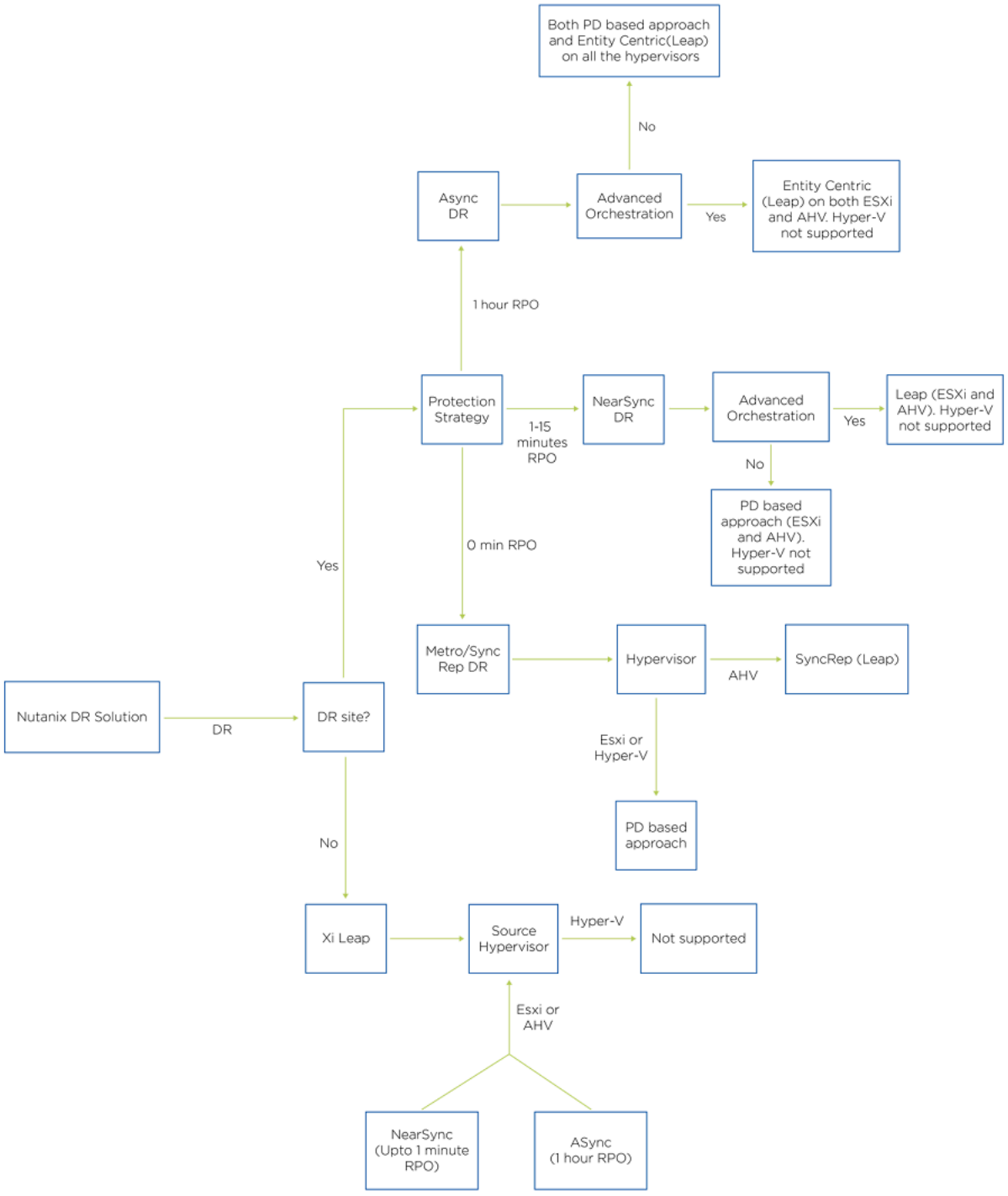

Figure 20: Disaster Recovery Decision Flowchart

1. Nutanix Metro Availability
   a. Pros: Can provide zero data loss (RPO 0), minimal to zero recovery time (RTO 0), and RTO 0 during a disaster recovery avoidance event.
   b. Cons: Requires stretched layer 2 networking of some form between edge sites or within the Edge site, requires a Nutanix Cloud Infrastructure Ultimate license, is only supported with ESXi, and is highly dependent on the network connectivity being resilient and providing a high enough throughput to accommodate application change rates.
   c. See the Metro Availability best practice guide on the Nutanix Support Portal for more information.
2. Synchronous replication
   a. Pros: Available with Nutanix AHV or VMware vSphere. Protects workloads with RPO 0. Used when you can't provide spanned L2 across datacenters but still need to provide RPO 0 to applications.
   b. Cons: Requires high bandwidth and low latency, requires a Nutanix Cloud Infrastructure Ultimate license or an advanced replication add-on license, only supported with Nutanix Disaster Recovery (previously Leap), and requires that the source cluster have more than three nodes.
3. NearSync replication
   a. Pros: Provides an RPO between 1 and 15 minutes to storage-intensive workloads over limited connectivity.
   b. Cons: Capacity of each SSD in the cluster must be at least 1.2 TB, only supported with Nutanix Disaster Recovery, requires a Nutanix AOS Enterprise license or a Leap Advanced Add-On license, and requires that the source cluster have more than three nodes.
4. Asynchronous replication
   a. Pros: Provides an RPO of one hour or more to applications using low-bandwidth, high-latency datacenter links.
   b. Cons: Requires a Nutanix AOS Enterprise license or Leap Advanced Add-On License and requires that the source cluster have more than three nodes.

General BCDR Best Practices and Requirements

- Keep as few VMs per protection domain as possible in your design.
  - If you're using asynchronous replication, configure at most 200 VMs per protection domain.
  - If you're using NearSync replication, configure at most 10 VMs per protection domain.
- Group VMs with similar RPO requirements in the same protection domain.
  - Node size and capacity can limit the possible RPO for the VMs running on it. See the Resource Requirements Supporting Snapshot Frequency (Asynchronous, NearSync and Metro) section of the Data Protection and Recovery with Prism Element guide for details.
- Put NearSync VMs in their own protection domains (NearSync can only have one schedule).
- Limit consistency groups to fewer than 10 VMs.
- Only use consistency groups for applications with a shared state, such as database replicas.
- Only add one VM to any consistency group that uses application-consistent snapshots.
- Configure snapshot intervals to be shorter than your desired RPO to allow for failure and recovery without manual intervention. The interval should be half your desired RPO including data transmission timing.
- Configure snapshot schedules to retain the minimum number of snapshots needed to meet your retention policy.
- If you're using synchronous replication, create one container on the source and one on the destination with the same name.
- If you're using synchronous replication, ensure that the two participating clusters are within 5 ms round trip time (RTT) of each other.

- Use Metro Availability for business-critical applications that require zero data loss (RPO 0), minimal to zero recovery time (RTO 0), or RTO 0 during a disaster recovery avoidance event.
  - Metro Availability is only supported with ESXi.

For additional details, see the Data Protection and Recovery with Prism Element guide on the Nutanix Support Portal.

6. Optional Nutanix Products and Services

Nutanix offers optional products and services that you can choose to incorporate when implementing enterprise edge solutions on the Nutanix platform. These options can help you quickly satisfy unique organizational requirements.

- Nutanix Cloud Manager: NCM Starter and Pro licenses add to the base capabilities of Nutanix Prism to provide monitoring, analytics, and automation capabilities from a single management interface.
- Nutanix Cloud Clusters (NC2): NC2 extends the Nutanix stack to the public cloud, creating a single management domain that spans private and public clouds and removes the friction from multicloud operations. For more information, see the Nutanix Cloud Clusters on AWS tech note and the Disaster Recovery with Nutanix Cloud Clusters (NC2) on AWS tech note.
- Flow Network Security: Microsegmentation and other capabilities offered by Flow Network Security further enhance the security of your Nutanix operations.
- Nutanix Files: Files enables Nutanix clusters with full-featured file services capabilities, eliminating the need for standalone NAS appliances or file servers.
- Nutanix Objects: Objects enables a Nutanix cluster to provide S3-compatible, software-defined, scale-out object stores for a variety of use cases, eliminating the need for standalone object storage while supporting applications that rely on object stores with an on-premises alternative.
- Nutanix Mine: Mine is an integrated backup solution that combines the benefits of the Nutanix architecture with the capabilities of proven backup vendors.

- NCM Self-Service: NCM Self-Service (previously Calm) provides application-level orchestration and lifecycle management, simplifying application management and enabling self-service.
- Karbon: Karbon integrates certified Kubernetes with the Nutanix operating environment, enabling Nutanix clusters to support both VM- and container-based applications while simplifying Kubernetes deployment and management.

7. Conclusion

The Nutanix Cloud Platform remains a common platform for both traditional datacenters and enterprise edge sites. This document is intended to demonstrate valuable methods and practices that organizations, both large and small, can use to implement Nutanix solutions to solve their IT and business problems. There is no one-size-fits-all solution for enterprise edge deployments; this document is only intended to provide customers with suggested best practices they can use to design and evolve their private, hybrid, and multicloud solutions.

For more information on any details of this document, please visit our website at www.nutanix.com or reach out to our sales team. You can also contact one of the many global support phone numbers listed on our website. For feedback or questions, please contact us using the Nutanix NEXT Community forums.

8. Appendix

Best Practices for Enterprise Edge Sites

High-Level Design Best Practices

- Logically group the manageability, replication, and BCDR components of your edge sites in a single region or availability zone.
- Use Prism Central as the primary control plane.
  - If you use ESXi, you also need VMware vCenter.
- Use the simplified block-and-pod architecture detailed in the Block and Pod Architecture section of the Nutanix Hybrid Cloud Reference Architecture.
  - Large-scale sites (more than four blocks with the expectation of additional blocks in the future) with the characteristics of a datacenter should reference the Nutanix Hybrid Cloud Reference Architecture.
- Build clusters with one or three nodes.
  - If you need to be able to scale your edge solution beyond two nodes or add storage-only nodes to the cluster, you should implement a three-node cluster.
- If you use them, deploy storage-only nodes in pairs.
  - You can only use storage-only nodes in clusters with three or more nodes.
- Define an SOP that stipulates how to scale one- or two-node clusters in your design, as this step isn't a function of the core AOS and doesn't have a workflow.

Software Best Practices

Use the following software versions:

- Prism Central: pc.2020.7

- Nutanix AOS 5.20
- Nutanix AHV 5.15
- VMware vSphere 6.7 or 7.0
- VMware vCenter 6.7 or 7.0

Failure Domain Best Practices

- Design and deploy the management plane to be highly available.
- At a minimum, design to meet the availability requirements of the managed workload or service with the highest uptime requirement.
- Confine workload domain to a single edge site.
- Confine workloads domain to a defined security zone.
- Ensure that the API gateway is always available because other third-party integrations (such as the backup vendor integration) may rely on it.
- Configure built-in RBAC to restrict access to management platform resources.
- Design your management plane to reside in the region or location where your edge clusters reside.
- Use redundant power from two different power supplies.
- Use redundant top-of-rack switches.
- Ensure redundant upstream connectivity to top-of-rack switches from each Nutanix node.
- Ensure redundant connectivity from edge to central datacenter.
- Use Nutanix scale-out architecture and data protection capabilities to replicate data to a second Nutanix cluster if one cluster fails.
- Deploy the application across multiple clusters.
- Use an independent and redundant cooling system for each datacenter room.

- Ensure redundant networking connectivity from different providers between datacenter buildings.
- Ensure redundant internet connectivity from different providers.
- Where possible, use multiple buildings and server rooms in separate buildings or campuses so that the failure of a single server room doesn't affect production. Distribute multicomponent services equally across datacenter server rooms.
- Ensure strict SLA compliance and keep well-documented resources and processes to help keep MTTR short.

General Networking Best Practices

- Configure IPMI, iLO, or iDRAC server-facing ports in the management network as access ports and don't use VLAN trunking for these ports. Restrict access to this critical management network.
- Never use these switches for Nutanix deployments:
  - Cisco Nexus 2000 (Fabric Extender)
  - 10 Gbps expansion cards in a 1 Gbps access switch
- In small edge sites running a handful of VMs, use a simple network topology consisting of a pair of top-of-rack switches interconnected to a firewall and router.
- In medium and larger deployments with multiple racks, use a leaf-spine topology because it's easy to scale, achieves high performance with low latency, and provides resilience.
- Ensure that every Nutanix node in a cluster is in the same layer 2 broadcast domain and shares the same IP subnet.
- You can populate nodes designed exclusively for edge or ROBO with 1 Gbps networking for one- and two-node clusters with latency spikes of approximately 6–8 ms.
  - Limit your cluster size to at most eight nodes if you plan on using 1 Gbps networking.

- Ensure that when there are two or more nodes in a cluster, each layer of the network topology is highly available and tolerates individual device failures.
  - Avoid configurations or technologies that don't maintain system availability during single-device outages or upgrades, such as stacked switches.
- Ensure that there are no more than three switches between any two Nutanix nodes in the same cluster.
- Don't use WAN or remote links between Nutanix nodes in the same Nutanix cluster.
- Separate Nutanix CVM and hypervisor hosts in a dedicated VLAN that doesn't include any VM traffic.
- Don't place Nutanix nodes in the same Nutanix cluster if the stretched layer 2 network spans multiple edge sites, buildings, datacenters, or availability zones or if there's a remote link between the two locations.
  - Using a stretched layer 2 network over a layer 3 network is only acceptable when the Nutanix cluster is in the same switch fabric or aggregation layer, such as when a layer 2 network stretches between two racks in the same datacenter.
- Don't use features like block or rack awareness to stretch a Nutanix cluster between different physical sites.
- Configure adequate uplinks between switches or interswitch links for east-west storage traffic to minimize port-to-port oversubscription. For example, use multiple 40 Gbps uplinks (or interswitch links).
- Connect hosts using redundant links.
- Configure switch ports facing Nutanix servers as spanning tree portfast or edge to skip the listening and learning phases and prevent cluster outages caused by changes to spanning tree topology.
- Configure the CVM and hypervisor VLAN as native, or untagged, on server-facing switch ports. Newly added nodes use untagged traffic for discovery and work out of the box, reducing manual server configuration.

- Use tagged VLANs on the switch ports for all guest workloads to keep the workloads separate from each other and from the CVM and hypervisor network.
- Reduce network oversubscription to ensure as close to a one-to-one ratio as possible. Dropped network packets or a congested network immediately affect storage performance.

AHV Networking Best Practices

- Use only the default bridge, br0, with at least two of the fastest network uplink adapters of the same speed to simplify the design.
  - › Converge the management, storage, and workload traffic on this single pair of uplink adapters.
- Only add additional brN bridges when you need a connection to a separate physical network. For example, if the top-of-rack switch has two pairs of switches—one pair for storage and management and another pair for workload traffic—it makes sense to create another bridge, br1, and place the workloads on this bridge.
- Don't modify the configuration of any bridges inside the AHV host unless following an official Nutanix guide.
- To keep network configuration simple and reduce risk, use the standard 1,500-byte MTU in the hosts, CVMs, and workload VMs. Nutanix doesn't recommend jumbo frames unless specifically required by high-performance Nutanix Volumes iSCSI workloads or specific workload requirements.
  - › When switching from 1,500-byte frames to 9,000-byte frames, performance improvements are generally not significant unless the workload uses the maximum network bandwidth for read traffic. For more information on when to use jumbo frames, see the Nutanix Volumes best practice guide and AHV Networking best practice guide.
- Connect at least one 10 Gbps or faster NIC to each tor switch to maintain high availability if one switch is lost.
- Use NICs from the same vendor within a bond to ensure compatibility and prevent undesired failover behavior.

- Use VLANs to separate logical networks. Physical hosts have a limited number of network ports and each port adds complexity. You can separate traffic logically without numerous physical ports.
- To simplify the design, use active-backup uplink load balancing.

vSphere Networking Best Practices

- Don't modify vSwitchNutanix.
- Convert vSwitch0 to the virtual distributed switch (VDS) following the instructions in Migrate from a Standard Switch to a Distributed Switch. Converting to the VDS allows you to centrally manage networking for all hosts, instead of configuring each host network individually. The VDS also enables advanced networking functions such as load-based teaming, LACP, and traffic shaping.
- To simplify the design, connect at least two of the fastest adapters of the same speed to vSwitch0 and use the Route Based on Physical NIC Load load balancing method to ensure that traffic is balanced between uplink adapters.
  - Connect these adapters to two separate top-of-rack switches to ensure redundancy.
- Connect at least one 10 Gbps or faster NIC to each top-of-rack switch to maintain high availability if one switch is lost.
- Don't add more vSwitches unless you need to connect to another physical network to meet security or workload requirements.
- All CVM storage, hypervisor host, and workload traffic should flow through vSwitch0, using VLANs to separate the workload traffic and all other traffic.
- Use the default 1,500-byte frame size on all uplinks unless there is a specific performance or application requirement that justifies 9,000-byte jumbo frames.

Virtualization Best Practices

- If you're using the disaster recovery runbook automation feature, you need to deploy separate vCenter and Prism Central instances in each region or availability zone.
- Enable HA to automatically restart VMs in case of node failure.
- Enable DRS with the default automation level so that it can move VMs as needed to ensure optimal performance.

  Note: Moving VMs may temporarily affect Nutanix data locality.

- Configure DAS.IGNOREINSUFFICIENTHBDATASTORE if one Nutanix container is presented to the ESXi hosts to eliminate false positives when a cluster uses a single datastore.
- Disable automation level, HA, and DRS for all CVMs. HA doesn't need to reserve resources for CVMs because they're bound to a single node and can't be restarted elsewhere in the cluster.
- Set Host failures cluster tolerates to 1 for replication factor 2 and 2 for replication factor 3. These settings ensure that the proper amount of resources are automatically reserved for the cluster.
- Set the host isolation response to Power off and restart VMs to ensure that VMs are moved to a healthy host and continue to function.
- Set the host isolation response to Leave powered on for CVMs. You don't want CVMs to be turned off if there's a transient network disruption.
- Disable Storage I/O Control. If Storage I/O Control is enabled, it can cause storage unavailability, unnecessary lock files, and complications with Metro Availability.

Management Best Practices

- Deploy a scaled-out Prism Central instance for edge sites to simplify capacity planning, platform lifecycle management, and virtual networking management, and to reduce management overhead.

- Deploy one instance of vCenter HA for all workloads and configure it with Platform Services Controller embedded.
- Configure at least three NTP servers (NTP standard recommendation is five to detect rogue time sources) and ensure that they're accessible at all solution layers, including AOS, AHV, and Prism Central, plus vCenter and ESXi if you're using vSphere.
- Use the same NTP servers for all infrastructure components.
- Don't use an AD domain controller as an NTP source.
- If you're in a dark site with no internet connectivity, use a switch or GPS time source.
- Configure NTP sources specific to the region where the edge clusters reside.
- Configuring at least two DNS servers and make them accessible at all layers (AOS, Prism Central, ESXi, vCenter, and network switches) to ensure that components can reliably resolve addresses at all times.
- Configure DNS sources specific to the region where the edge clusters reside.
- Use RBAC for all the consoles and element managers and use AD as the directory service for user and group accounts.
- Where possible, use AD groups to assign privileges for easier operations.
- Deploy a robust logging platform that meets your security requirements.
- Forward all infrastructure logs to the centralized log repository.
- Store the logs in a different cluster, or location, from where you collect them. This measure protects the logs in case of a catastrophic cluster failure, ensuring they can later be used for forensics.
- Ensure that AD sites and services are set up to define a replication and authentication topology in the event that connectivity to the datacenter is lost.

Security Best Practices

Authentication

- Maintain as few user and group management systems as possible. A centrally managed authentication point is preferred to many separately managed systems.
- You should at least take advantage of the external LDAP support provided by Nutanix components.
- Use AD authentication for user and server accounts to ensure that user activity logged for auditing purposes and account security is configured and maintained from a single centralized solution.
  - › You must have a highly available AD infrastructure and network connection to the datacenter.
- Use an SSL or TLS connection to AD to eliminate cleartext exchanges on the network.

Certificates

- Protect all consumer-facing components with certificates signed by a trusted certificate authority to provide an extra layer of security and prevent meddler-in-the-middle (MITM) attacks.
  - › You can use internally or externally signed certificates based on your consumer classification and the service the specific component provides.

Cluster Lockdown

- Don't use Nutanix or vSphere cluster lockdown (feature that lets you enforce SSH access to CVMs and host using key pairs instead of passwords) unless you require passwordless communication.

Hardening

- Enable Advanced Intrusion Detection Environment (AIDE) for the CVM and AHV. AIDE performs checksum verification for all statis binaries and libraries for improved security.
- Enable stack traces for cluster issues for AHV or ESXi and CVM.

- Enforce a complex password policy (at least 15 characters long with at least 8 different characters) for the hypervisor and CVM.
- Enable a banner for AHV or ESXi and the CVM that retrieves a specific sign-in message via SSH.
- Enforce only SNMPv3 on the CVM.
- Configure SCMA to run hourly to capture unacceptable configuration drift.
- ESXi only: Stop unused services and close unused firewall ports to limit the attack surface.

  Note: Ensure that you don't stop a service or close a firewall port that's required by Nutanix, such as SSH and NFS.

- Generally, we recommend following the Hardening Controller VM section of the AOS Security Guide. If you're using AHV, follow the Hardening AHV section of the AOS Security Guide.

Internet-Facing Services

- Use multiple internet connections (active-backup) and ISP-provided denial-of-service (DoS) and distributed-denial-of-service (DDoS) filtering to help mitigate the potential effects of a DoS or DDoS attack.

Logging

- Configure and maintain a single centralized logging solution for auditing purposes and account security.
- Send log files to a highly available syslog infrastructure.
- At least one of the individual targets of the highly available logging infrastructure should run outside the virtual infrastructure itself so that if the virtual infrastructure is compromised, forensic investigations can access logs that might not be available on the cluster itself.
- Include data from all Nutanix modules and components in logging using the error-log level and ensure that they're searchable. See the Configuring the Remote Syslog Server Settings section of the Acropolis Advanced Administration Guide for more information.

- Use default ESXi logging levels, log rotation, and log file sizes.
- If you have additional security and reliability requirements, use TCP for log transport. Otherwise, use the default syslog protocol, UDP.
- Use port 514 (defined port in syslog RFC) for logging.

Network Segmentation

- To protect Nutanix CVM and hypervisor traffic, place them together in their own dedicated VLAN, separate from other VM traffic. This configuration applies to all hosts in a cluster or single node.
- Place out-of-band management on a separate VLAN or physical network to provide additional security.
- We recommend configuring the CVM and hypervisor host VLAN as a native, or untagged, VLAN on the connected switch ports to allow easy node addition and cluster expansion.
- Don't segment Nutanix storage and replication traffic, or iSCSI Volumes traffic, on separate interfaces (VLAN or physical) unless additional segmentation is required by mandatory security policy or the use of separate physical networks.

Role-Based Access Control (RBAC)

- Use a least-privilege and separation-of-duties approach when assigning RBAC permissions to ensure that each group or individual user has just enough permissions to perform their duties. Use predefined roles or create new roles as needed.
- Configure RBAC at the Prism Central level because it provides the overlying management construct. This configuration ensures that your least-privilege configuration stays in place and avoids common mistakes that occur when RBAC is configured at multiple different levels.
- Align RBAC structure and default plus custom roles with your company requirements.

Data-at-Rest Encryption

- Keep management traffic, including storage traffic, on a separate network.
- Don't use storage encryption.
- Don't use a key management server (KMS).

Microsegmentation and Firewall

- Use Flow Network Security (previously Flow Microsegmentation) to comply with regulatory and business-specific compliance policies such as PCI-DSS, HIPAA, and NIST.
- Use a Palo Alto Networks VM-Series firewall with Flow Network Security to provide advanced threat and vulnerability detection for layers 4–7.

Operations Best Practices

- Purchase the Nutanix Cloud Manager (NCM) Starter license, which gives you access to analytics, capacity planning, custom dashboards, and playbooks for Prism Central.
- Review capacity planning monthly.
- Perform updates after hours for performance- or migration-sensitive applications to reduce the load on the system.
- Use the current Long-Term Support (LTS) branch unless you require a specific new feature for your design.
- Update to the next maintenance version four weeks after release and to the current patch version two weeks after release to keep updates relatively small.
- Maintain a preproduction environment to test any necessary changes to firmware, software, or hardware before implementing the change in production. Preproduction environments can help you avoid the cost of outages.
- Configure alerts and alert policies in Prism Central, not Prism Element, to create consistency across multiple clusters and reduce the effort involved in

making multiple changes. Prism Central–based alert policies also allow you to detect anomalies.

- Use SMTP for alert transmission to create consistency across multiple clusters, reduce the effort involved in making multiple changes, enable anomaly detection, and offer more options for delivery.

BCDR Best Practices

- Keep as few VMs per protection domain as possible in your design.
  - If you're using asynchronous replication, configure at most 200 VMs per protection domain.
  - If you're using NearSync replication, configure at most 10 VMs per protection domain.
- Group VMs with similar RPO requirements in the same protection domain.
  - Node size and capacity can limit the possible RPO for the VMs running on it. See the Resource Requirements Supporting Snapshot Frequency (Asynchronous, NearSync and Metro) section of the Data Protection and Recovery with Prism Element guide for details.
- Put NearSync VMs in their own protection domains (NearSync can only have one schedule).
- Limit consistency groups to fewer than 10 VMs.
- Only use consistency groups for applications with a shared state, such as database replicas.
- Only add one VM to any consistency group that uses application-consistent snapshots.
- Configure snapshot intervals to be shorter than your desired RPO to allow for failure and recovery without manual intervention. The interval should be half your desired RPO including data transmission timing.
- Configure snapshot schedules to retain the minimum number of snapshots needed to meet your retention policy.

- If you're using synchronous replication, create one container on the source and one on the destination with the same name.
- If you're using synchronous replication, ensure that the two participating clusters are within 5 ms round trip time (RTT) of each other.
- Use Metro Availability for business-critical applications that require zero data loss (RPO 0), minimal to zero recovery time (RTO 0), or RTO 0 during a disaster recovery avoidance event.
  - › Metro Availability is only supported with ESXi.

References

1. AHV best practice guide
2. AHV Networking best practice guide
3. Data Efficiency tech note
4. Data Protection and Disaster Recovery best practices guide
5. Disaster Recovery with Nutanix Cloud Clusters (NC2) on AWS tech note
6. Flow Network Security tech note
7. Hybrid Cloud Reference Architecture
8. Information Security tech note
9. Nutanix AHV Virtualization tech note
10. Nutanix DRaaS Security tech note
11. Nutanix Era Security tech note
12. Nutanix Hardware Compatibility Lists
13. Nutanix Volumes best practice guide
14. Palo Alto Networks VM-Series Firewalls on Xi tech note
15. Physical Networking best practice guide
16. Prism tech note
17. ROBO Deployment and Operations
18. Security Advisories
19. Virtual Machine High Availability tech note
20. VMware vSphere best practice guide
21. VMware vSphere Documentation

About Nutanix

Nutanix is a global leader in cloud software and a pioneer in hyperconverged infrastructure solutions, making clouds invisible and freeing customers to focus on their business outcomes. Organizations around the world use Nutanix software to leverage a single platform to manage any app at any location for their hybrid multicloud environments. Learn more at www.nutanix.com or follow us on Twitter @nutanix.

List of Figures