v1.1 | August 2022 | TN-2030

TECH NOTE

Red Hat OpenShift on Nutanix

NUTANIX
YOUR ENTERPRISE CLOUD

Copyright

Contents

1. Executive Summary 5

2. Introduction 7
- Audience 7
- Purpose 7
- Document Version History 7

3. Nutanix Cloud Platform Overview 8

4. Red Hat OpenShift Overview 9

5. Architecture 11
- Control Plane 11
- Infrastructure Nodes 11
- Compute Nodes 12

6. Sizing 13
- Compute 13
- Network 14
- Storage 14

7. Deployment Type 16

8. Virtual Machine Placement and ADS 17
- VM-VM Antiaffinity Rules for Control Plane and Infrastructure Nodes 17
- VM-VM Antiaffinity Rules for Compute Nodes 18

9. Storage Integration 19
- Prerequisites 19
- Nutanix CSI Operator 20
- Red Hat OpenShift Monitoring 20

OpenShift Image Registry....20

10. Networking....22

11. GPU....24

12. Backup and Restore....25
Control Plane....25
Applications....25

13. Conclusion....27

14. Appendix....28
References....28

About Nutanix....29

List of Figures....30

1. Executive Summary

As organizations go cloud native and build and deploy containerized applications at scale, they run into big challenges with building a hybrid cloud–ready Kubernetes stack on-premises. Most cloud-native infrastructure solutions don't offer the scalability and resilience that enterprise Kubernetes demands, nor do they seamlessly integrate key infrastructure layers or simplify life cycle management and developer tooling. Given the complex challenges involved with operationalizing an enterprise Kubernetes environment, cloud-native enterprises demand best-in-class infrastructure solutions that are vendor-certified and jointly supported.

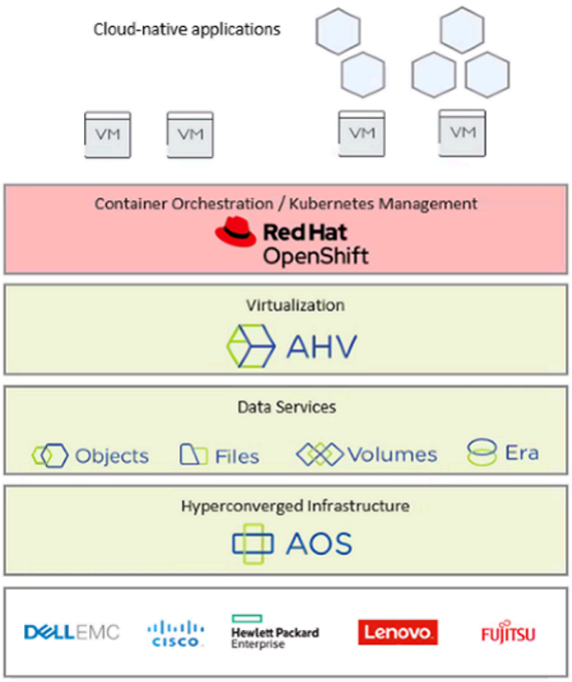

Figure 1: Nutanix Cloud Platform with Red Hat OpenShift

Nutanix and Red Hat's strategic partnership offers enterprise customers a best-in-class solution for building, scaling, and managing cloud-native applications on-premises and in a hybrid cloud. This certified and jointly supported solution

features Red Hat OpenShift—the leading enterprise DevOps and container management platform—powered by Nutanix Cloud Platform, which seamlessly integrates Nutanix (an industry-leading hyperconverged infrastructure), AHV (a hypervisor certified for Red Hat Enterprise Linux), and feature-rich data services. Red Hat OpenShift on Nutanix provides a resilient and scalable infrastructure and a cloud-native application platform, enabling organizations to deliver innovative solutions, drive competitive advantage, and meet customer expectations.

2. Introduction

Audience

This tech note is part of the Nutanix Solutions Library. We wrote it for individuals responsible for designing, building, managing, and supporting Red Hat OpenShift on Nutanix infrastructures. Readers should be familiar with Nutanix AOS, Nutanix Files, Nutanix Objects, and Red Hat OpenShift.

Purpose

In this document, we cover the following topics:

- Planning and deploying OpenShift on Nutanix AHV.
- Integrating OpenShift into Nutanix Unified Storage.
- Appropriate strategies for integrating with an existing deployment.

Unless otherwise stated, the solution described in this document is valid on all supported AOS releases.

Document Version History

Version Number	Published	Notes
1.0	May 2022	Original publication.
1.1	August 2022	Updated Deployment Type section.

3. Nutanix Cloud Platform Overview

Based on the industry-leading hyperconverged infrastructure (HCI) that Nutanix pioneered, Nutanix Cloud Platform (NCP) is the ideal infrastructure choice for Kubernetes and enterprise hybrid cloud platforms like Red Hat OpenShift. We built Nutanix Cloud Infrastructure (NCI) to handle the changing demands of web-scale containerized applications and to linearly scale performance and capacity without limit, making it the ideal infrastructure for supporting enterprise cloud-native workloads. NCP is resilient from the ground up: nodes are self-healing and upgrade nondisruptively.

NCP also integrates a complete set of data services to handle the demands of stateful virtualized and containerized applications. Nutanix's full-featured Container Storage Interface (CSI) deploys with every Kubernetes cluster and natively integrates with Nutanix Files (simple, flexible, and intelligent scale-out file storage) and Nutanix Volumes (block storage with built-in space efficiency). In addition, Nutanix Objects provides a simple, scalable, and secure S3-compatible storage tier for workloads requiring object storage.

With Nutanix Database Service (NDB) Nutanix provides enterprise database administrators with a simple and effective means of deploying highly available databases and managing them with a single control plane across public, private, and hybrid cloud deployments. Databases can be run in VMs and be easily consumed by workloads running on Red Hat OpenShift.

NCP is a benchmark for simplicity that enables customers to streamline their Day 1 and Day 2 operations, easily support tier-1 enterprise workloads at scale, and serve the resource and services needs of software developers.

4. Red Hat OpenShift Overview

Red Hat OpenShift gives organizations a single platform for application innovation that lets them operate consistently and innovate continuously. It supports every application and can span every cloud so that organizations can be ready for today and build for the future. Red Hat OpenShift includes the Red Hat Enterprise Linux (RHEL) operating system, over-the-air updates, container runtime, networking, ingress, monitoring, logging, container registry, and authentication and authorization solutions. These components are tested together for unified operations on a complete Kubernetes platform spanning every cloud.

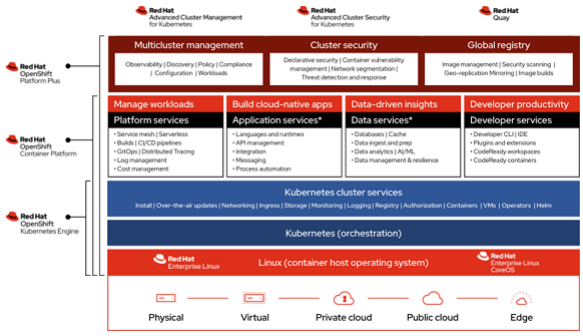

Figure 2: Red Hat OpenShift

As the preferred choice for enterprise full-stack Kubernetes solutions on NCP, Red Hat OpenShift supports diverse workloads across the hybrid cloud. Whether you're modernizing existing applications, building new cloud-native applications, integrating data analytics and AI or ML, or using software from independent software vendors (ISVs), Red Hat OpenShift provides the flexibility to choose the applications that make the most sense for your business.

With Red Hat OpenShift on NCP, IT operations teams can focus on managing and supporting new cloud-native applications and modernizing existing workloads to add more value to the business, rather than managing the complex underlying infrastructure.

For more information about Red Hat OpenShift, visit the Red Hat OpenShift product page.

5. Architecture

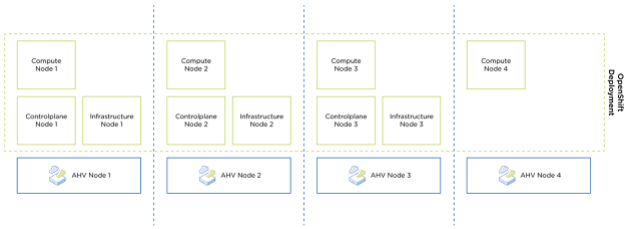

Figure 3: OpenShift Components Distributed across a Nutanix Cluster

Control Plane

The control plane, which is composed of control plane nodes, manages the OpenShift Container Platform cluster. The control plane nodes manage workloads on the infrastructure and compute nodes.

Infrastructure Nodes

Infrastructure nodes allow customers to isolate infrastructure workloads to avoid incurring billing costs against subscription counts and to separate maintenance and management.

These nodes host only infrastructure components, such as the default router, the integrated container image registry, and the components for cluster metrics and monitoring. These infrastructure machines are not counted toward the total number of subscriptions that are required to run the environment.

In a production deployment, we recommend deploying at least three nodes to hold infrastructure components. Both OpenShift Logging and Red Hat OpenShift Service Mesh deploy Elasticsearch, which requires three instances to be installed on different nodes.

Compute Nodes

The OpenShift compute nodes run containerized applications created and deployed by developers. An OpenShift application node can run RHCOS or RHEL.

The size and number of compute nodes required depends on application needs. It's also possible to have different sets of nodes with special purposes, such as nodes with different CPU/RAM ratios or GPU-enabled nodes.

6. Sizing

For information on minimum requirements and sizing recommendations, view Red Hat's documentation:

- Minimum resource requirements for cluster installation
- How to plan your environment according to application requirements

Compute

Sizing an OCP cluster is a multistep process that involves both application and infrastructure needs. Determine the workload requirements at the application level, the overhead for the necessary basic services running on each compute node, and a high availability (HA) reserve in case of compute node failure. This calculation gives you the total CPU (in cores) and main memory (in MB) requirements.

You can calculate the required number of compute nodes based on the previous values and the number of available physical hosts. In the absence of any additional information, start with three nodes—one per physical node—vertically scaled with respect for sizing guidelines like NUMA architecture.

If pod density becomes an issue (for example, if the application pods are small and there are many instances, you might exceed the limit of 500 pods per node), scale out in increments (one additional compute node per physical node).

> Note: To reach a max of 500 Pods, you may need to configure custom kubelet settings and use the host-prefix /22.

If the number of compute nodes exceeds the number of the physical hosts and you provide multiple compute nodes per host, you risk creating an imbalance where the failure of a physical node can leave the cluster without enough resources to restore the lost pods. In this instance, AHV helps by providing the HA reserve to restart the compute nodes on other AHV nodes. Note that the

HA reserve doesn't ensure distribution of a deployment within OpenShift across different AHV nodes (for more details, see the Recommendations section).

The recommended size for the control plane is based on the number of compute nodes, pods, and namespaces.

For a single OpenShift cluster, we don't recommend using CPU oversubscription. In general, Kubernetes has its own scheduler to place pods on nodes, which might create CPU hot spots if CPU oversubscription is used.

From a Red Hat licensing perspective, you should avoid CPU overprovisioning because compute nodes use a vCPU-based license. Overprovisioning limits the utility of that license.

You can allow a certain amount of overprovisioning if you have multiple independent OpenShift clusters running on the same AHV cluster and use some of these OpenShift clusters for low-priority test environments.

Network

There may be additional sizing factors to consider involving the number of compute and infrastructure nodes, like the amount of ingress traffic.

For infrastructure nodes, start with three nodes. Scale up as needed to accommodate application ingress traffic, but account for physical limitations, too. For example, if the application needs 60 Gbps of ingress traffic, but each node only has a 10 GbE NIC, then the NICs become a bottleneck.

Storage

The requirements of both your OpenShift environment and your applications drive your storage needs.

All components of the OpenShift cluster use a local attached disk for storing data.

The control planes use an etcd database for all configuration data, which expects very low latency write operations. The data persists on these disks.

The local disk of a compute node provides read access to the container images and to the ephemeral storage of a running container. Write operations use this ephemeral storage heavily when running build processes within the OpenShift environment.

To provide best performance for both the control plane and compute nodes, choose write-intensive SSDs or NVMe for the underlying hardware. Hybrid clusters are also fine, just as long as write-intensive SSDs are used for the performance layer.

Persistent storage for the applications will be provided by Nutanix CSI, using Nutanix Volumes and/or Nutanix Files, or by Nutanix Objects if the applications can consume S3-compatible storage. Size these storage services based on the specified requirements of the applications.

Services running within the OpenShift environment also consume persistent storage by storing metrics and logging data.

7. Deployment Type

There are different ways to deploy OpenShift on Nutanix HCI depending on the hypervisor you choose. This document covers planning and deploying OpenShift on Nutanix AHV. For both supported hypervisors (AHV and ESXi), you can choose from several supported installation methods.

You can integrate all installation methods into automation workflows for deployment in the block and pod architecture.

With the release of OpenShift 4.11, Nutanix HCI is also a supported target for the installer-provisioned installation (IPI) process. The OpenShift install tool and the Nutanix Controller Virtual Machine (CVM) provide all necessary infrastructure.

These features simplify Day 1 and Day 2 operations like scaling your compute machine sets or easily creating new ones for infrastructure nodes.

The official OCP documentation should be the ultimate source for any installation information. Additionally, Nutanix-maintained documentation provides several post-install steps, which are helpful for additional configuration of storage and internal registry components.

8. Virtual Machine Placement and ADS

Acropolis Dynamic Scheduling (ADS) is responsible for examining the load on various components for hotspots and calculating what migrations can be done to balance the load without causing additional hotspots. ADS monitors the following resources:

- Total CPU usage of each of the guest VMs.
- Storage controller (Stargate) CPU usage per VM or iSCSI target.

Creating dedicated VM-VM antiaffinity groups for control plane, infrastructure, and compute nodes places the machines on different physical hosts by best effort to accommodate a host failure and allows you to use all one-click-features and updates.

VM-VM Antiaffinity Rules for Control Plane and Infrastructure Nodes

Sign in to the CVM with SSH to create the VM groups and configure antiaffinity. If you are running multiple OpenShift deployments, create separate groups per deployment:

```
acli vm_group.create controlplane ocp1
acli vm_group.create infra-ocp1
acli vm_group.antiaffinity_set controlplane-ocp1
acli vm_group.antiaffinity_set infra-ocp1
```

Assign your VMs to the corresponding groups (make sure to use the correct VM names used in your deployment):

```
acli vm_group.add_vms controlplane-ocp1 vm_list=controlplane-*
acli vm_group.add_vms infra-ocp1 vm_list=infra-*
```


VM-VM Antiaffinity Rules for Compute Nodes

Within a single group, there shouldn't be more VMs than the cluster node count. As soon as the number of compute nodes exceeds the number of AHV hosts, the additional VMs would no longer be distributed evenly across the hosts.

If you want to use VM-VM antiaffinity rules in this case, you may need to create additional groups.

Be aware that running multiple compute nodes on the same AHV node will require additional application configuration to ensure that a deployment, stateful set, or replication controller will not place all pods on the same physical node. An easy way to ensure this is to use a replica count higher than the number of compute nodes per AHV node. As another solution, you may introduce labels, taints, and tolerations on the compute nodes and applications.

9. Storage Integration

Prerequisites

Nutanix Volumes

To access Nutanix Volumes from the CSI driver, you need to configure the data services IP on your Nutanix cluster.

Nutanix Files

To create rwx persistent volume claims, you need Nutanix Files as your back-end NFS storage. Follow the Nutanix Files documentation for detailed information on how to deploy it.

Nutanix Objects

Follow the Nutanix Objects documentation for detailed information on how to deploy an objects store on Nutanix.

If necessary, download the Nutanix Objects Certificate Authority (CA) from Prism Central, and use the Objects GUI to add it as the trusted CA in OpenShift. You will find instructions in the corresponding sections.

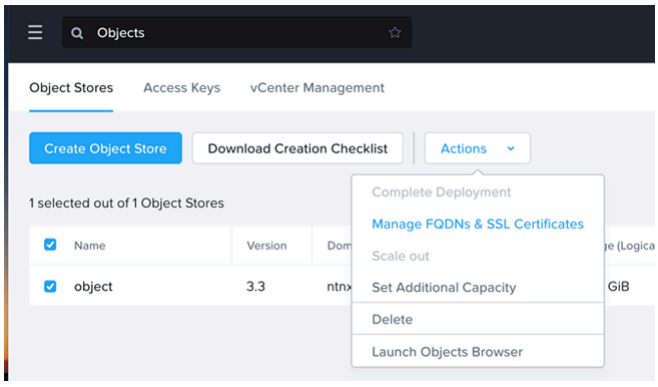

Figure 4: Managing Certificates from the Objects GUI

Nutanix CSI Operator

The Nutanix CSI Operator for Kubernetes packages, deploys, manages, and upgrades the Nutanix CSI Driver on Kubernetes and OpenShift for dynamic provisioning of persistent volumes on the Nutanix Cloud Platform.

The CSI Driver for Kubernetes leverages Nutanix Volumes and Nutanix Files to provide scalable and persistent storage for stateful applications.

With Files storage classes, applications on multiple pods can access the same storage and have the benefit of multi-pod read and write access.

To install the Nutanix CSI Operator, follow the instructions in the CSI Operator documentation.

Red Hat OpenShift Monitoring

To retain your metrics data, it's necessary to add persistent storage for Prometheus and AlertManager. Add a Volume Claim Template in the `cluster-monitoring-config` for `prometheusK8s` and `AlertmanagerMain` components.

Edit the `cluster-monitoring-config` ConfigMap object in the `openshift-monitoring` project:

```
oc -n openshift-monitoring edit configmap cluster-monitoring-config
```

Add your PVC configuration for the component under data/config.yaml:

```
data:
  config.yaml: |
    <component>:
      volumeClaimTemplate:
       spec:
         storageClassName: <storageclass-name>
         resources:
           requests:
             storage: <ammount of storage>
```

OpenShift Image Registry

OpenShift provides a highly available, OCI-compliant image registry. It's preferred by Red Hat to use an S3-compliant storage as a back end, which is

provided by Nutanix Objects. Based on individual customer needs, it is also possible to use Persistent Volumes (rwx) provided by Nutanix Files.

After deploying OpenShift on Nutanix, the image registry is not configured and is set to unmanaged.

To use Nutanix Objects as your back-end storage, complete the following steps:

- Create a ConfigMap from the downloaded PEM file.

```
oc create configmap object-ca \
--from-file=ca-bundle.crt=object-ca.pem \
-n openshift-config
```

- Assign the ConfigMap to the global proxy-settings.

```
oc patch proxy/cluster \
--type=merge \
     --patch='{"spec":{"trustedCA":{"name":"object-ca"}}}'
```

- Create a secret containing your bucket credentials.

```
oc create secret generic image-registry-private-configuration-user \
    --from-literal=REGISTRY_STORAGE_S3_ACCESSKEY=my_access_key \
    --from-literal=REGISTRY_STORAGE_S3_SECRETKEY=my_secret_key \
    --namespace openshift-image-registry
```

- Patch the image registry to use the bucket.

```
oc patch configs.imageregistry.operator.openshift.io/cluster \
    --type='json' \
    --patch='[
{"op": "remove", "path": "/spec/storage" },
{"op": "add", "path": "/spec/storage", "value":
{"s3":
{"bucket": "image-registry-bucket",
"regionEndpoint": "https://my.objects.endpoint",
"encrypt": false,
"region": "us-east-1"}}}]'
```

- Enable the image registry in OpenShift.

```
oc patch configs.imageregistry.operator.openshift.io cluster --type merge --patch
 '{"spec":{"managementState":"Managed"}}'
```


10. Networking

To add a supplementary layer of hardening to the OCP cluster, use Flow Network Security. Flow Network Security is part of NCI and introduces microsegmentation and service-chaining to VM traffic granularly based on categories. You can only use Flow Network Security with Nutanix AHV. If you use a different hypervisor, you should use its corresponding microsegmentation solution.

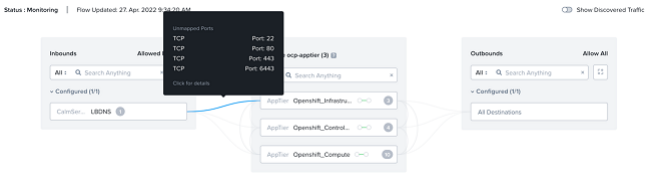

Figure 5: Visual Representation of Microsegmentation

A basic strategy for microsegmentation at a VM level is to allow only necessary communication between the OpenShift layers as documented in the OpenShift network connectivity requirements.

The following table lists the necessary categories you must create to implement this strategy.

Table: Flow Network Security Categories for Red Hat OpenShift

Category Type	Category Name	Assigned VMs
AppType	Openshift	All OCP VMS
AppTier	Openshift_Controlplane	Control plane VMs
AppTier	Openshift_Infrastructure	Infrastructure VMs
AppTier	Openshift_Compute	Compute VMs
User-defined	Your_Loadbalancer_Name	Load Balancer

Assign the OpenShift nodes to their corresponding AppTier category, which you use to control, visualize, and log the data flow.

Based on OpenShift's network connectivity requirements, create an application security policy in Flow with the load balancer category as an allowed inbound service and the three AppTier layers based on the created categories. Then, assign rules for inbound traffic and traffic between AppTiers.

11. GPU

You can configure NVIDIA GPUs as passthrough devices or vGPUs. After you assign the GPU resource to a set of compute nodes, follow the NVIDIA GPU Operator on OpenShift documentation. Essentially, passthrough mode runs the node feature discovery operator to label the compute nodes with all the necessary information. The NVIDIA GPU Operator uses these labels to assign the necessary drivers to the compute nodes.

Using vGPU involves further steps, like creating a dedicated driver image. After configuring the GPU on AHV, use NVIDIA's GPU Operator on OpenShift documentation to create a dedicated driver image if needed.

For information about supported NVIDIA GPUs on AHV, refer to the AHV Administration Guide. You can also download an AHV-compatible NVIDIA driver from the Nutanix Support Portal.

12. Backup and Restore

Control Plane

The OpenShift control plane should be included in existing VM-level backup strategies. For dedicated backups of an etcd database, see the OpenShift control plane backup and restore operations documentation.

Applications

To back up and restore applications running on Red Hat OpenShift, use the OpenShift API for Data Protection (OADP).

OADP backs up and restores Kubernetes resources and internal images at the granularity of a namespace by using Velero. Nutanix CSI offers snapshot capabilities, which can be leveraged by OADP to back up and restore persistent volumes (PVs).

After you've deployed the OADP Operator, create a credential file:

```
cat << EOF > ./credentials-velero
[default]
aws_access_key_id=my_access_key
aws_secret_access_key=my_secret_key
EOF
oc create secret generic cloud-credentials -n openshift-adp --from-file
 cloud=credentials-velero
```

Next, create a Nutanix CSI Snapshot-Class and label it for use with OADP:

```
apiVersion: snapshot.storage.k8s.io/v1beta1
kind: VolumeSnapshotClass
metadata:
  name: nutanix-snapshot-class
  labels:
    velero.io/csi-volumesnapshot-class: "true"
driver: csi.nutanix.com
parameters:
  storageType: NutanixVolumes
  csi.storage.k8s.io/snapshotter-secret-name: ntnx-secret
  csi.storage.k8s.io/snapshotter-secret-namespace: ntnx-system
deletionPolicy: Delete
```

Lastly, create an OADP Application:


```
apiVersion: oadp.openshift.io/v1alpha1
kind: DataProtectionApplication
metadata:
  name: oadp-ntnx
  namespace: openshift-adp
spec:
  configuration:
    velero:
      defaultPlugins:
        - openshift
        - aws
        - csi
    featureFlags:
    - EnableCSI
    restic:
      enable: false
  backupLocations:
    - name: default
      velero:
        provider: aws
        default: true
        objectStorage:
          bucket: oadp
          prefix: velero
          caCert: <base64 encoded Objects CA>
        config:
          insecureSkipTLSVerify: "true"
          region: us-east-1
          s3ForcePathStyle: "true"
          s3Url: https://object.ntnxlab.local
        credential:
          key: cloud
          name: cloud-credentials
```


13. Conclusion

The Nutanix Cloud Platform and AHV provide a powerful foundation for the proven capabilities of Red Hat OpenShift. Together they're a flexible, reliable, and economical solution for running containerized workloads in your datacenter. Nutanix streamlines and enhances both storage infrastructure configuration and overall deployment, helping you build an on-premises, cloud-native infrastructure.

For feedback or questions, contact Nutanix using the NEXT Community forums.

14. Appendix

References

1. CNCF definition of cloud native technology
2. Red Hat OpenShift product page
3. Minimum resource requirements for OCP cluster installation
4. How to plan your OCP environment according to application requirements
5. Red Hat OpenShift Container Platform on Nutanix AOS
6. OCP documentation
7. Adding an ISCSI Data Services IP Address (Cluster Details)
8. Nutanix Files documentation
9. Nutanix Objects documentation
10. CSI Operator documentation
11. OpenShift network connectivity requirements
12. AHV Administration Guide
13. AHV-compatible NVIDIA drivers
14. OpenShift control plane backup and restore operations

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

Figure 1: Nutanix Cloud Platform with Red Hat OpenShift....5

Figure 2: Red Hat OpenShift....9

Figure 3: OpenShift Components Distributed across a Nutanix Cluster....11

Figure 4: Managing Certificates from the Objects GUI....19

Figure 5: Visual Representation of Microsegmentation....22