Best Practices Recommendation

NetApp Solutions

NetApp
September 13, 2024

Table of Contents

Best Practices Recommendation

Best practices recommendations for VMs in Red Hat OpenShift Virtualization

Author: Banu Sundhar, NetApp

This section describes the different factors you should consider when deploying new VMs or when importing existing VMs from a VMware environment into OpenShift Virtualization on OpenShift Container Platform.

VM performance

When creating a new VM in OpenShift Virtualization, you need to consider the access pattern along with performance (IOPs and throughput) requirements of the workload that will run on the VM. This will influence the number of VMs you will need to run on the OpenShift Virtualization in an OpenShift Container platform and the type of storage that you need to use for the VM disks.

The type of storage you want to choose for your VM disks are influenced by the following factors:

- The protocol access you need for data access of your workloads
- The access modes you need (RWO vs RWX)
- The performance characteristics you need for your workloads

See the Storage Configuration section for more details.

High Availability of VM workloads

OpenShift Virtualization supports Live migrations of a VM. Live migration allows a running Virtual Machine Instance (VMI) to move to another node without interrupting the workload. Migration can be helpful for a smooth transition during cluster upgrades or any time a node needs to be drained for maintenance or configuration changes.
Live migration requires the use of a shared storage solution that provides ReadWriteMany (RWX) access mode. The VM disks should be backed by storage option that provides RWX access mode. OpenShift Virtualization will check that a VMI is **live migratable** and if so the **evictionStrategy** will be set to **LiveMigrate**.
See About live migration section in Red Hat documentation for details.

It is important that you use a driver that supports **RWX** access mode.
See the Storage Configuration section for more details about what drivers support RWX access mode.

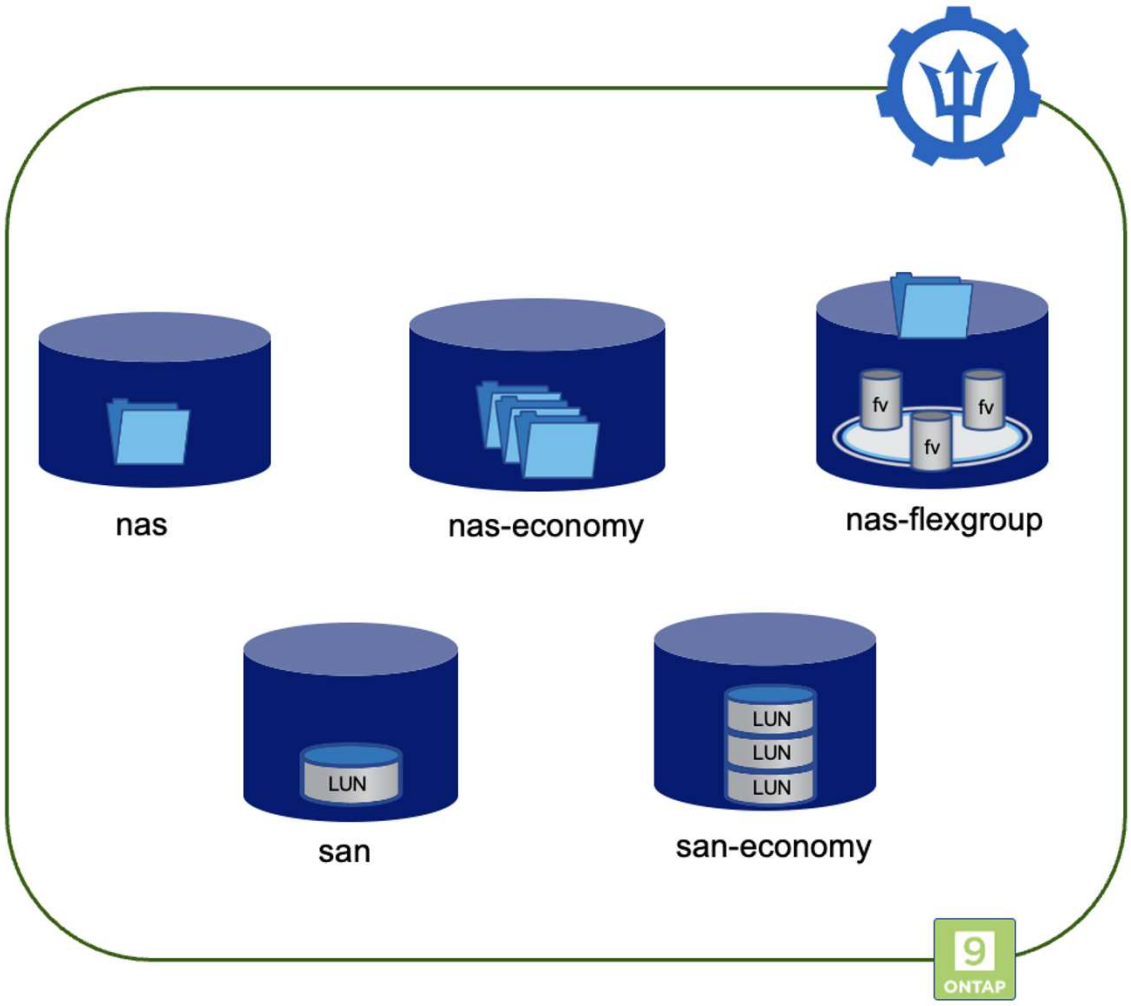
Storage Configuration

Trident CSI provisioner provides several drivers (nas, nas-economy, nas-flexgroup, san and san-economy) for provisioning storage backed by NetApp storage options.
Protocols used:
* nas drivers use NAS protocols (NFS and SMB)
* san drivers use iSCSI or NVMe/TCP protocol

The following can help you decide how you want the storage configuration based on the workload requirements and storage utilization.

- nas driver creates one persistent volume (PV) on one FlexVolume.
- nas-economy driver creates one PV on a qtree on a shared FlexVolume. (one FlexVolume for every 200 PVs, configurable between 50 and 300)
- nas-flexgroup driver creates on one PV on one FlexGroup
- san driver creates one PV on LUN on a dedicated FlexVolume
- san-economy driver creates one PV on LUN on shared FlexVolume (one FlexVolume for every 100 PVs, configurable between 50 and 200)

The following diagram illustrates this.



Also, the access modes supported by the drivers differ.

ONTAP nas drivers support

- Filesystem access and RWO, ROX, RWX, RWOP access modes.

ONTAP san drivers support raw block as well as filesystem modes.

- In the raw block mode, it can support RWO, ROX, RWX, RWOP access modes.
- In the filesystem mode, only RWO, RWOP access modes are permitted.

Live migration of OpenShift Virtualization VMs require the disks to have RWX access modes. So, it is important that you choose nas drivers or san drivers in raw block volume mode to create PVCs and PVs backed by ONTAP.

Storage Configuration Best Practices

Dedicated Storage Virtual Machines (SVMs)

Storage Virtual Machines (SVMs) provide isolation and administrative separation between tenants on an ONTAP system. Dedicating an SVM to OpenShift containers and to OpenShift Virtualization VMs enables the delegation of privileges and enables applying best practices for limiting resource consumption.

Limit the maximum volume count on the SVM

To prevent Trident from consuming all the available volumes on the storage system, you should set a limit on the SVM. You can do this from the command line:

```
vserver modify -vserver <svm_name> -max-volumes <num_of_volumes>
```

The max-volumes value is the total volumes provisioned across all the nodes in the ONTAP cluster, and not on an individual ONTAP node. As a result, you might encounter some conditions where an ONTAP cluster node might have far more or less Trident provisioned volumes than another node. To avoid this, ensure that equal number of aggregates from each node in the cluster are assigned to the SVM used by Trident.

Limit the maximum size of volumes created by Trident

You can set a maximum volume size limit on a per SVM basis in ONTAP:

1. Create the SVM with the vserver create command and set the storage limit:

```
vserver create -vserver vserver_name -aggregate aggregate_name -rootvolume
root_volume_name -rootvolume-security-style {unix|ntfs|mixed} -storage
-limit value
```

1. To modify the storage limit on an existing SVM:

```
vserver modify -vserver vserver_name -storage-limit value -storage-limit
-threshold-alert percentage
```

Storage limits cannot be configured for any SVM that contains data protection volumes, volumes in a SnapMirror relationship, or in a MetroCluster configuration.

In addition to controlling the volume size at the storage array, you should also leverage Kubernetes capabilities.

1. To configure the maximum size for volumes that can be created by Trident, use the **limitVolumeSize** parameter in your backend.json definition.
2. To configure the maximum size for FlexVols used as pools for ontap-san-economy and ontap-nas-economy drivers, use the **limitVolumePoolSize** parameter in your backend.json definition.

Use SVM QOS policy

Apply Quality of service (QoS) policy to the SVM to limit the number of IOPS consumable by the Trident provisioned volumes. This helps to prevent workloads using Trident provisioned storage from affecting workloads outside of the Trident SVM.

ONTAP QoS policy groups provide QoS options for volumes and enable users to define the throughput ceiling for one or more workloads.
For more information about QoS policy groups, refer to ONTAP 9.15 QoS commands

Limit storage resource access to Kubernetes cluster members

Use Namespaces
Limiting access to the NFS volumes and iSCSI LUNs created by Trident is a critical component of the security posture for your Kubernetes deployment. Doing so prevents hosts that are not a part of the Kubernetes cluster from accessing the volumes and potentially modifying data unexpectedly.

Also, a process in a container can access storage mounted to the host, but which is not intended for the container. Using Namespaces to provide logical boundary for resources can avoid this issue. However,

It's important to understand that namespaces are the logical boundary for resources in Kubernetes. Thus, it is critical to ensure that namespaces are used to provide separation when appropriate. However, privileged containers run with substantially more host-level permissions than normal. So, disable this capability by using pod security policies.

Use a dedicated export policy
For OpenShift deployments which have dedicated infrastructure nodes or other nodes which are unable to schedule user applications, separate export policies should be used to further limit access to storage resources. This includes creating an export policy for services which are deployed to those infrastructure nodes (for example, the OpenShift Metrics and Logging services), and standard applications which are deployed to non-infrastructure nodes.

Trident can automatically create and manage export policies. This way, Trident limits access to the volumes it provisions to the nodes in the Kubernetes cluster and simplifies the addition/deletion of nodes.

But if you choose to create an export policy manually, then populate it with one or more export rules that process each node access request.

Disable showmount for the application SVM
A pod deployed to the Kubernetes cluster can issue the showmount -e command against the data LIF and receive a list of available mounts, including those which it does not have access to. To prevent this, disable the showmount feature using the following CLI:

```
vserver nfs modify -vserver <svm_name> -showmount disabled
```

For additional details about Best Practices for Storage Configuration and Trident usage, review Trident documentation

OpenShift Virtualization - Tuning & Scaling Guide

Red Hat has documented OpenShift Cluster Scaling Recommendations and limitations.

In addition, they have also documented OpenShift Virtualization tuning guide and Supported Limits for OpenShift Virtualization 4.x.

An active Red Hat subscription is required to access the above content.

The tuning guide contains information about many tuning parameters including:

- Tuning parameters to create many VMs at once or in large batches
- Live migration of VMs
- Configuring a dedicated network for live migration
- Customizing a VM template by including a workload type

The supported limits document the tested object maximums when running VMs on OpenShift

Virtual Machine Maximums including

- Max virtual CPUs per VM
- Max and min memory per VM
- Max Single disk size per VM
- Max number of hot pluggable disk per VM

Host Maximums including
* Simultaneous live migrations (per node and per cluster)

Cluster Maximums including
* Maximum number of defined VMs

Migrating VMs from VMware Environment

Details about migrating VMs from VMware environment can be found under Workflows > Red Hat OpenShift Virtualization with NetApp ONTAP

If you are migrating more than 10 VMs from an ESXi host in the same migration plan, you must increase the NFC service memory of the host. Otherwise, the migration will fail because the NFC service memory is limited to 10 parallel connections. For additional details see the Red Hat documentation: xref:./containers/ Increasing the NFC service memory of an ESXi host

Copyright information

Trademark information