High Performance KubeVirt in Action

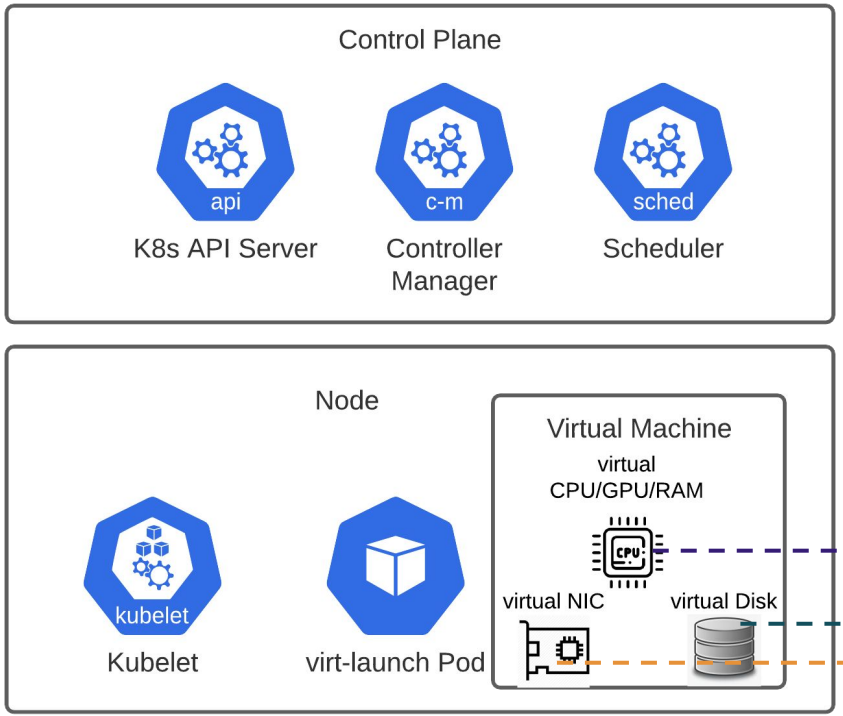
Huamin Chen
Red Hat
Twitter: root_fs
Github: rootfs

Marcin Franczyk
Kubermatic
Github: mfranczy

Outline

- KubeVirt Refresh
- Gardener Project Introduction
- Multus Network CNI for High Performance and Full Isolation
- Data Volume and Clone

KubeVirt Refresh

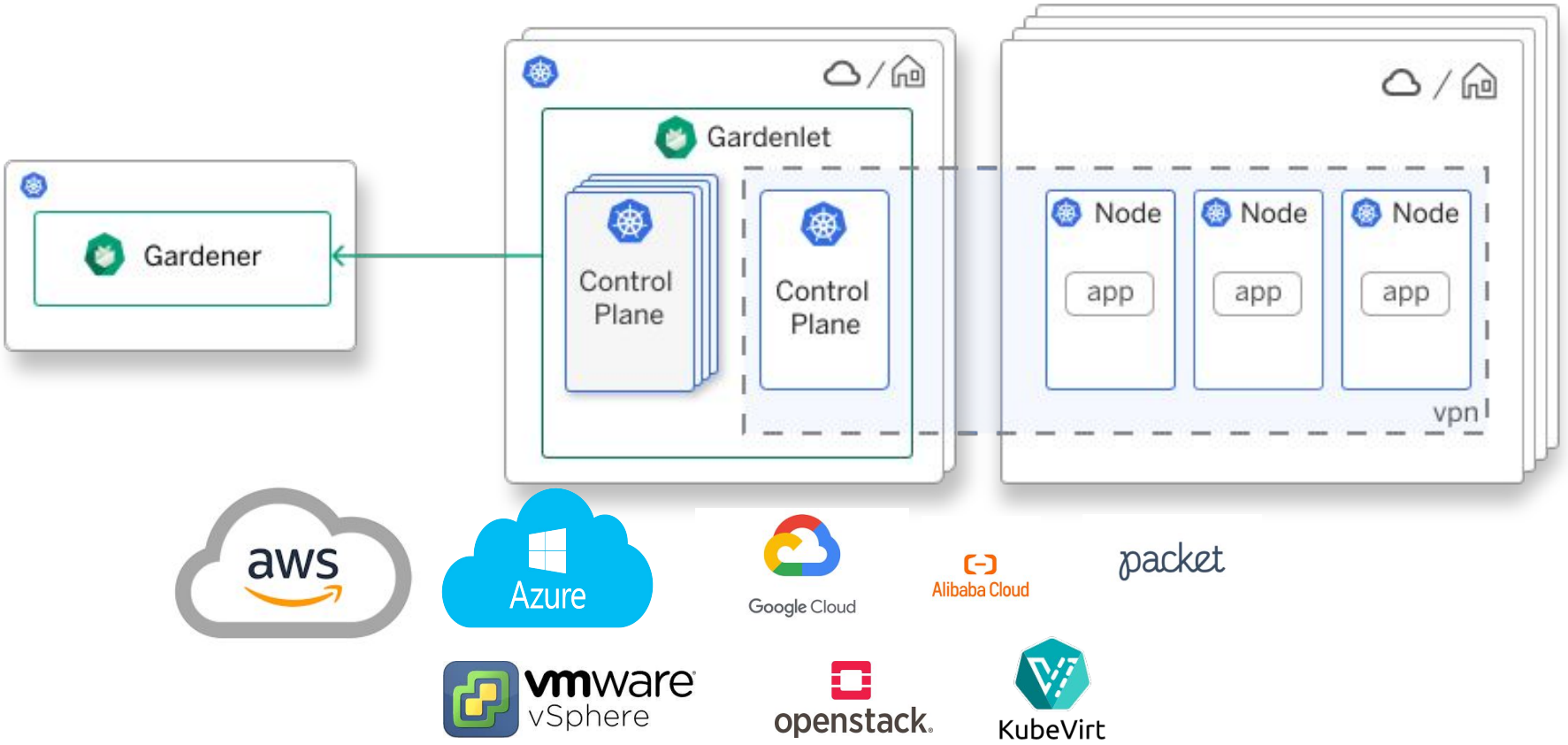



```
apiVersion: kubevirt.io/v1alpha3
kind: VirtualMachine
metadata:
  name: my-vm
spec:
  template:
   spec:
     domain:
      devices:
       interfaces:
       resources:
       disks:
       networks:
       volumes:
   dataVolumeTemplates:
     spec:
       pvc:
```

KubeVirt

KubeVirt provides declarative Virtual Machine lifecycle management on Kubernetes.

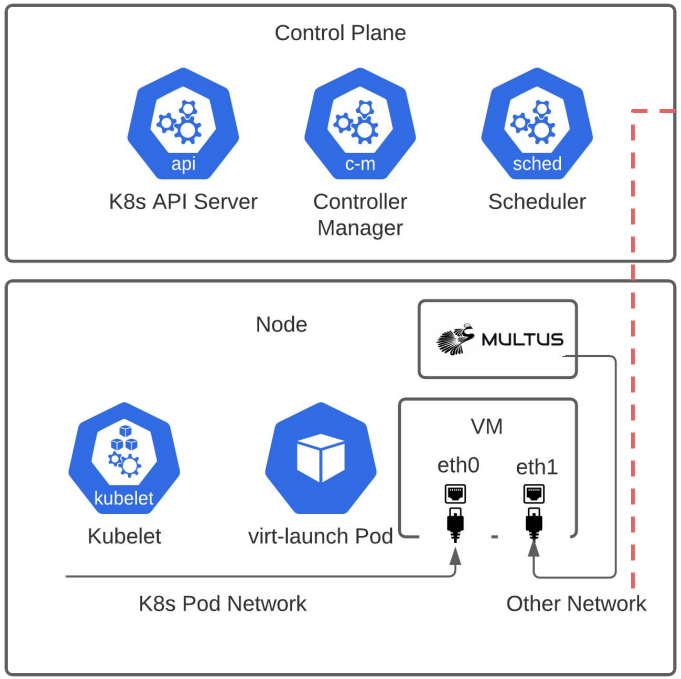
Gardener - Architecture



Multus for Network Extension and Isolation

```
apiVersion: k8s.cni.cncf.io/v1
kind: NetworkAttachmentDefinition
metadata:
  name: my-bridge
  namespace: my-ns
spec:
  config: '{
    "cniVersion": "0.4.0",
    "name": "my-bridge",
    "plugins": [
        { "name": "my-whereabouts",
          "type": "bridge",
          "bridge": "br1",
          "vlan": 1234,
          "ipam": {
            "type": "whereabouts",
            "range": "10.123.124.0/24",
            "routes": []
          }
        }
      ]
    }'
```

Multus for Network Extension



Multus is a meta CNI. It allows a Pod or VM to attach to other networks.

```yaml
apiVersion: k8s.cni.cncf.io/v1
kind: NetworkAttachmentDefinition
metadata:
  name: my-bridge
  namespace: my-ns
spec:
  config: '{
    "cniVersion": "0.4.0",
    "name": "my-bridge",
    "plugins": [
       {
          "name": "my-whereabouts",
          "type": "bridge",
          "bridge": "br1",
          "vlan": 1234,
          "ipam": {
            "type": "whereabouts",
            "range": "10.123.124.0/24",
            "routes": [
                { "dst": "0.0.0.0/0",
                  "gw" : "10.123.124.1" }
            ]}}]}'
```

A net-attach-def declares a Multus Plugin configuration. In this example, a Linux bridge using VLAN 1234 is created. It uses Whereabouts for IP address management.

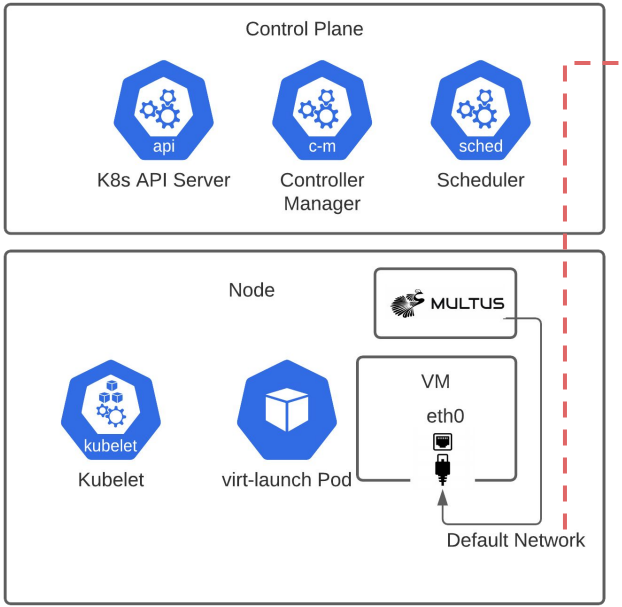
```yaml
apiVersion: kubevirt.io/v1alpha3
kind: VirtualMachine
metadata:
  name: my-vm
spec:
  template:
    spec:
      domain:
        devices:
          interfaces:
          - name: default
              masquerade: {}
          - bridge: {}
            name: other-net
      networks:
      - name: default
        pod: {}
      - multus:
          networkName:
            my-ns/my-bridge
          name: other-net
```

Pod Network

Multus Network

VM declaration references the net-attach-def using namespace/name notation.

Multus for Full Isolation




```yaml
apiVersion: k8s.cni.cncf.io/v1
kind: NetworkAttachmentDefinition
metadata:
  name: my-bridge
  namespace: my-ns
spec:
  config: '{
    "cniVersion": "0.4.0",
    "name": "my-bridge",
    "plugins": [
      {
        "name": "my-whereabouts",
        "type": "bridge",
        "bridge": "br1",
        "vlan": 1234,
        "ipam": {
          "type": "whereabouts",
          "range": "10.123.124.0/24",
          "routes": [
            { "dst": "0.0.0.0/0",
              "gw" : "10.123.124.1" }
          ]}}]}'
```

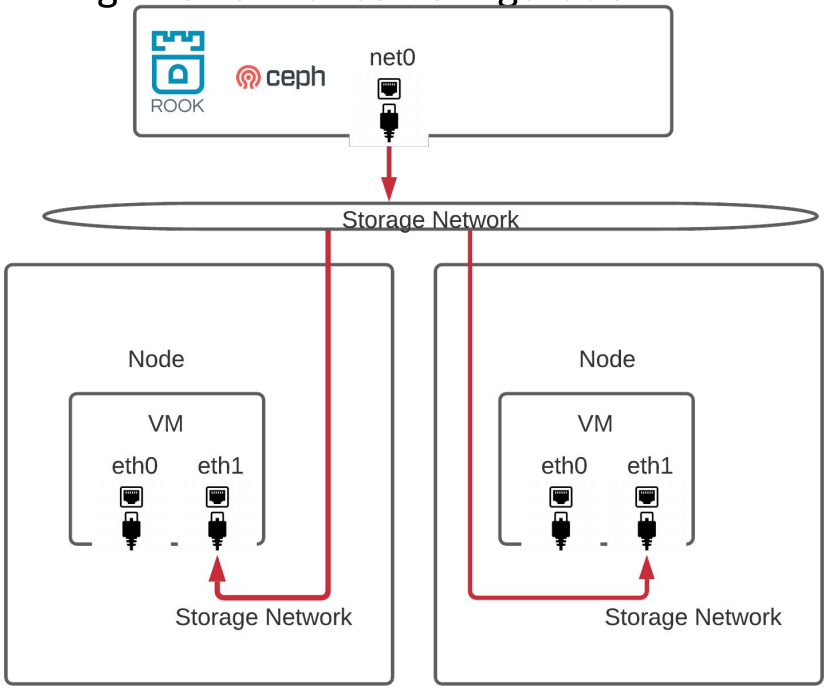
```yaml
apiVersion: kubevirt.io/v1alpha3
kind: VirtualMachine
metadata:
  name: my-vm
spec:
  template:
    spec:
      domain:
        devices:
          interfaces:
          - bridge: {}
            name: default-net
        networks:
        - multus:
            networkName:
              my-ns/my-bridge
          name: default-net
```

Multus Network

A VM can only uses Multus network, without attaching to the Kubernetes Pod Network, to achieve full isolation.

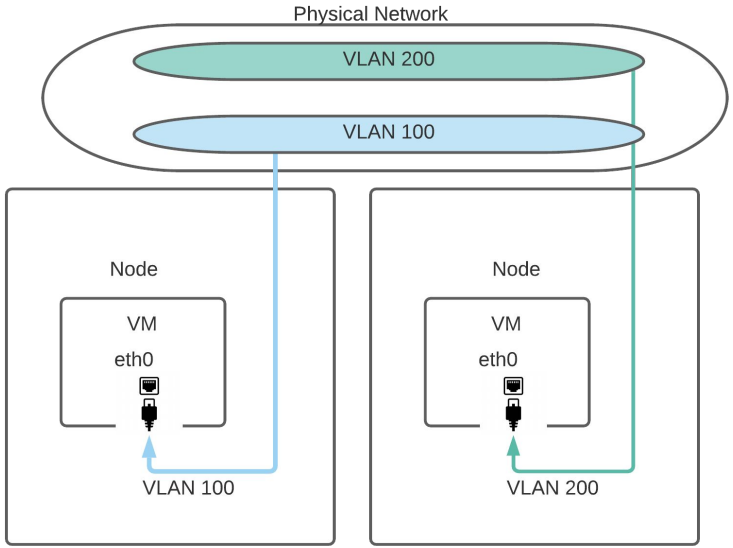
KubeVirt Use Cases

High Performance Configuration



VMs and Ceph attach to the same net-attach-def to achieve high performance.

Full Isolation Configuration

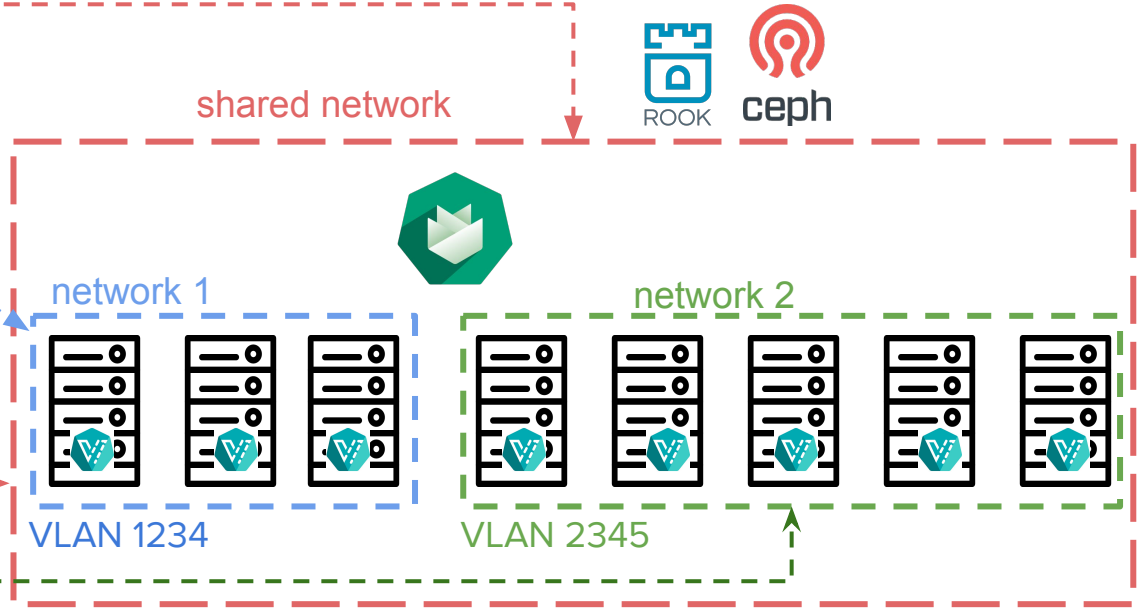


VMs use multus as default network and attach to different net-attach-def to achieve full isolation.

Gardener Shoot cluster network configuration

```yaml
apiVersion: core.gardener.cloud/v1beta1
kind: Shoot
infrastructureConfig:
    networks:
     sharedNetworks:
     - name: "ceph"
       name: ceph
       namespace: rook-ceph
     tenantNetworks:
     - name: "network1"
       default: true
       config: |-
       {
         "type": "bridge",
         "vlan": 1234,
       }
```

```yaml
apiVersion: core.gardener.cloud/v1beta1
kind: Shoot
infrastructureConfig:
    networks:
     sharedNetworks:
     - name: "ceph"
       name: ceph
       namespace: rook-ceph
     tenantNetworks :
     - name: "network2"
       default: true
       config: |-
       {
         "type": "bridge",
         "vlan": 2345,
       }
```



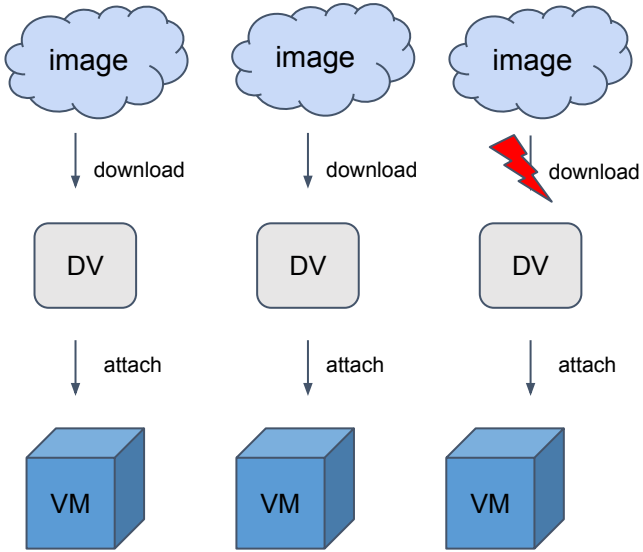
Multiple shoot clusters can access shared networks and create their own isolated tenant networks.

Data Volume and Clone

```yaml
apiVersion: cdi.kubevirt.io/v1alpha1
kind: DataVolume
metadata:
  name: my-cloned-dv
spec:
  source:
    pvc:
      name: my-source-dv
      namespace: my-ns
  pvc:
    accessModes:
    - ReadWriteOnce
    resources:
      requests:
        storage: 10Gi
      storageClassName: storage-provisioner
```

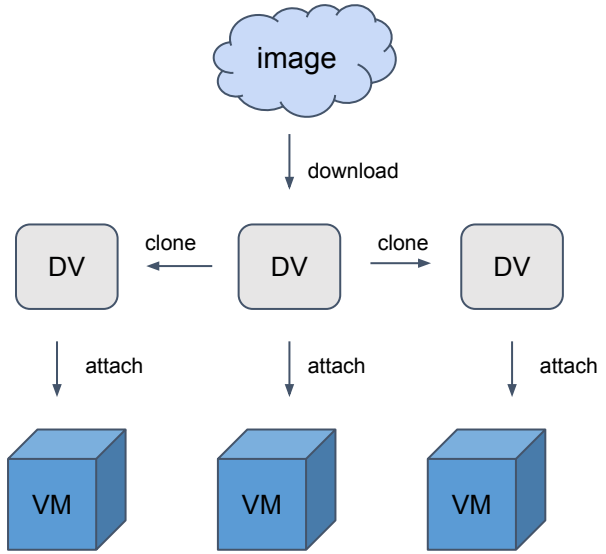
VM images management - Data Volume

Ad-hoc DataVolume



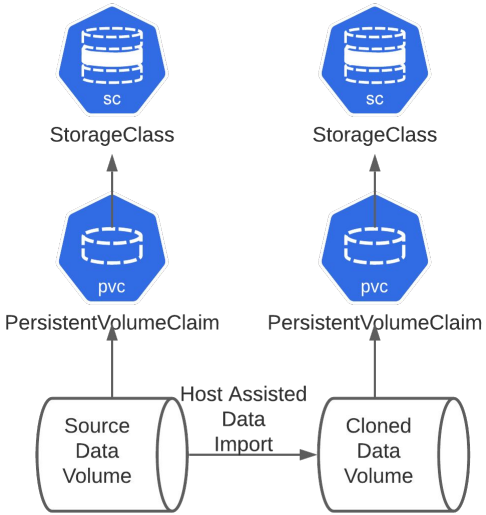
Downloading VM images to Data Volume on demand are prone to performance and reliability issues.

Pre-allocated DataVolume



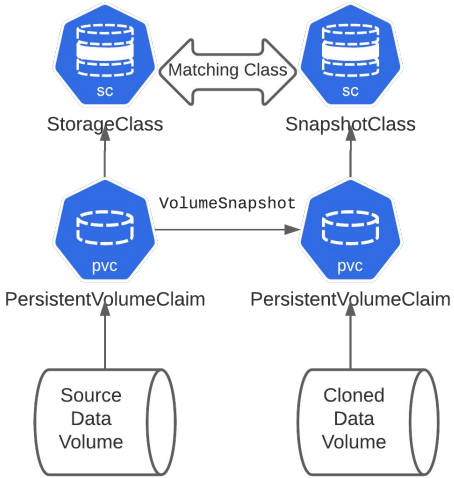
Pre-allocated Data Volumes only download images once, then are cloned to VMs upon creation, without having to re-downloading them, and thus reducing network latency.

Host Assisted Data Volume Clone



Host assisted Data Volume Clone can work in all cases but requires data copy.

Smart Clone

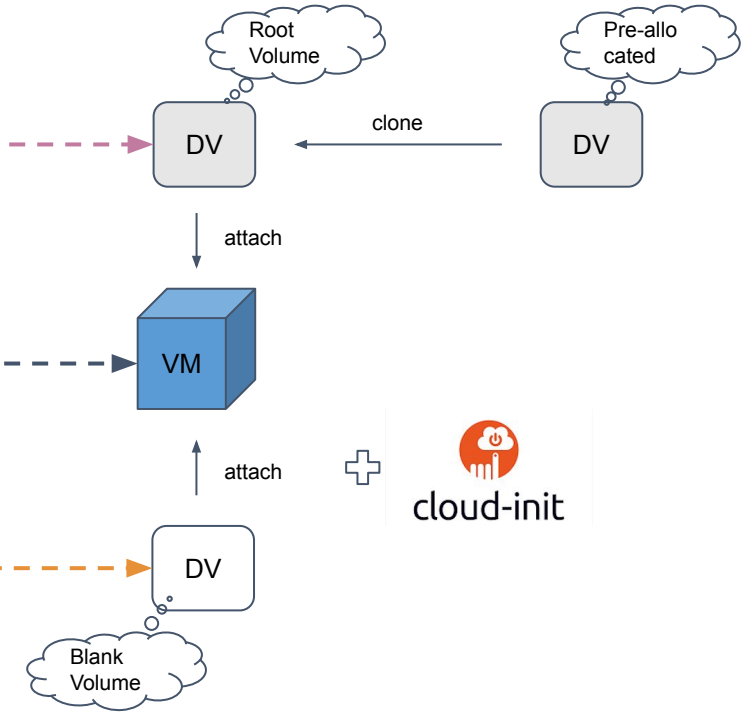


Smart Clone uses volume snapshot provided by Storage backends which is often implemented as Copy-on-Write (COW) and is thus more scalable.

Data Volume disks for shoot cluster nodes

```
apiVersion:
core.gardener.cloud/v1beta1
kind: Shoot
...
spec:
 provider:
   workers:
   - name: cpu-worker
     ...
     volume:
       size: 20Gi
       type: default
     dataVolumes:
     - name: volume-1
       size: 40Gi
       type: default
     ...
```

```
apiVersion: kubevirt.io/v1alpha3
kind: VirtualMachine
metadata:
  name: cpu-worker-asd12zxc
spec:
  dataVolumeTemplates:
  - metadata:
      name: root-disk
    spec:
      pvc:
        ...
            storage: 20Gi
        storageClassName: default
      source:
        pvc: ...
  - metadata:
      name: volume-1
    spec:
      pvc:
        ...
            storage: 40Gi
        storageClassName: default
      source:
        blank:{}
  template:
    spec:
      domain:
        devices:
          disks: ...
      volumes:
      - dataVolume:
          name: root-disk
        name: root-disk
      - dataVolume:
          name: volume-1
        name: volume-1
```



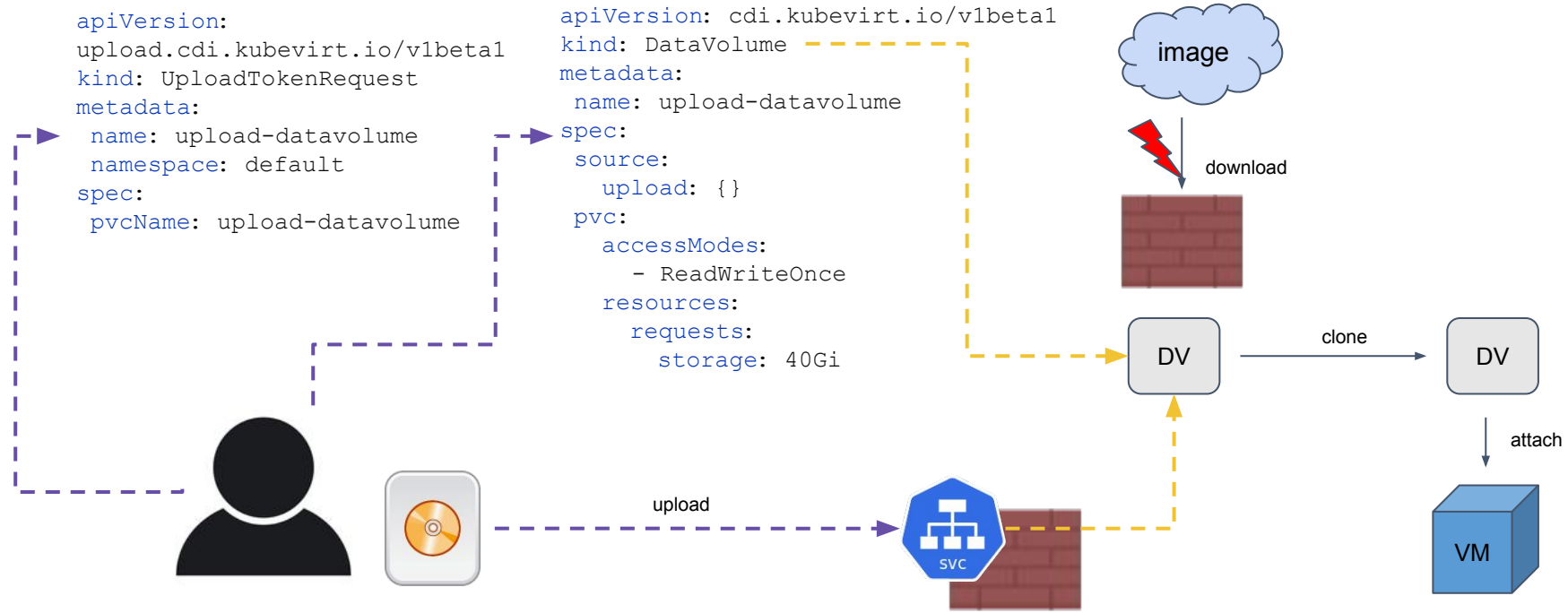
Shoot clusters can use pre-allocated Data Volume for fast root disks creation.

Data Volume provides an easy way of additional disks management.

Upload VM disk image to your cluster


```
apiVersion:
upload.cdi.kubevirt.io/v1beta1
kind: UploadTokenRequest
metadata:
 name: upload-datavolume
 namespace: default
spec:
 pvcName: upload-datavolume
```

```
apiVersion: cdi.kubevirt.io/v1beta1
kind: DataVolume
metadata:
 name: upload-datavolume
spec:
 source:
   upload: {}
 pvc:
   accessModes:
     - ReadWriteOnce
   resources:
     requests:
       storage: 40Gi
```



CDI upload helps to provision pre-allocated Data Volumes in highly isolated environments.


HELM
KV
KEEP CLOUD NATIVE
KubeCon
CloudNativeCon
North America 2020
Virtual
EVERYWHERE