

Kubernetes使用RBD

Subtitle

2022/10/05

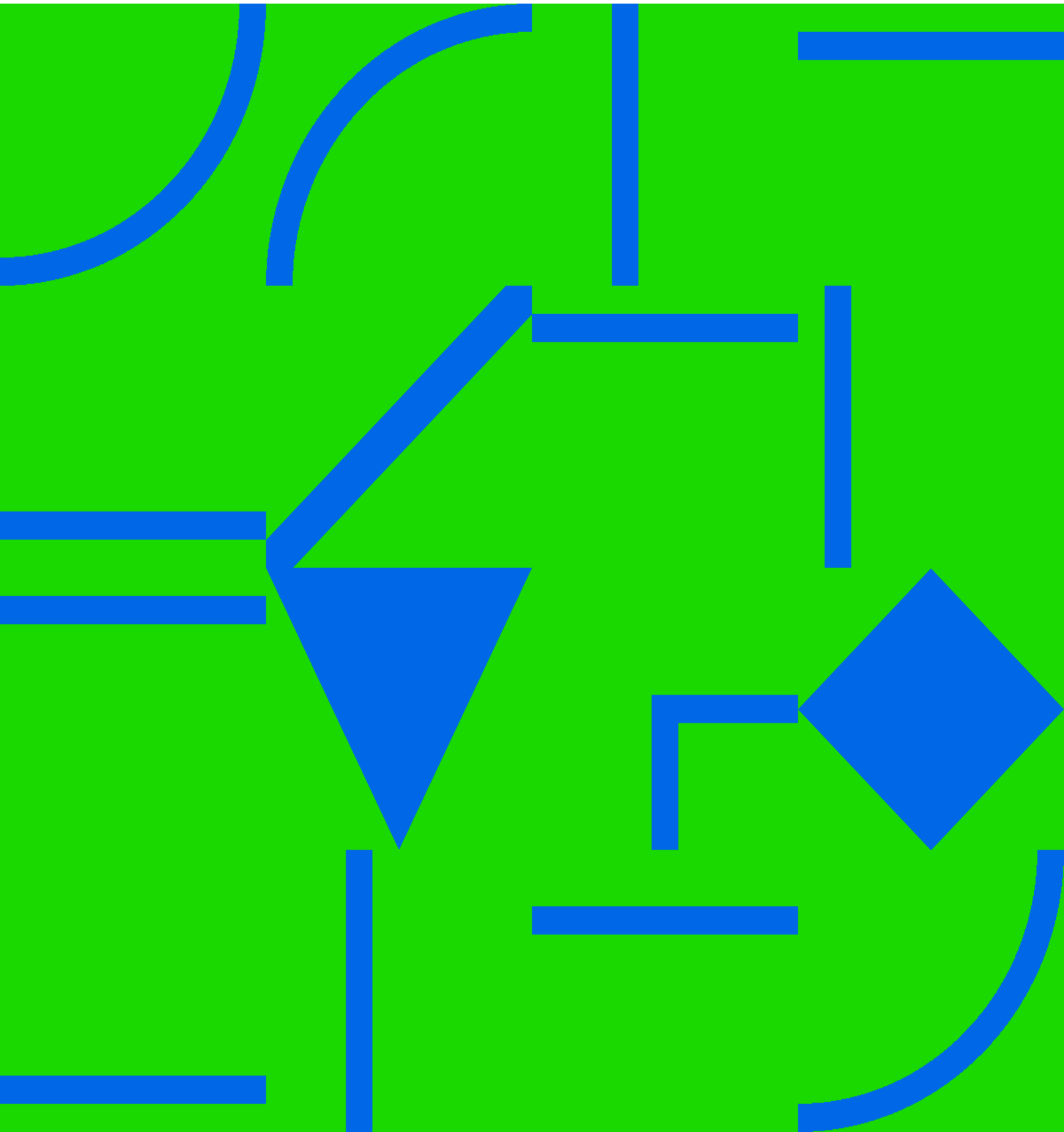


Table of Contents

Kubernetes使用RBD

前提条件

准备pool

使用流程

ceph准备pool和用户

创建 secret

创建 StorageClass

创建 PVC

引用 PVC

生命周期管理

更改pv回收策略

更改storageclass回收策略

问题

pv扩容报错

auth: unable to find a keyring

1

1

1

3

3

3

4

4

4

5

5

5

5

6

6

Kubernetes使用RBD

参考资料：

- <https://ieevee.com/tech/2018/05/16/k8s-rbd.html>
- <https://ieevee.com/tech/2019/04/14/pvc-resize.html>
- <https://www.cnblogs.com/ltxdzh/p/9173570.html>
- <https://amito.me/2018/Pools-and-Placement-Groups-in-Ceph/>
- <http://www.xuxiaopang.com/2016/11/17/exp-how-pg-affect-data-distribution/>

前提条件

- k8s node上需要 安装 ceph-common
- 内核版本须升级（centos 7.2.1511版本默认内核无法映射 rbd，升级到 7.6.1810 版本，内核 3.10.0-957 以上）

准备pool

pg_num计算器：<https://ceph.com/pgcalc/>

计划对象存储，rbd 和 fs 共用一个集群，开发环境和生成环境区分pool 为 kube.rbd, kube.fs, kube.dev.rbd, kube.dev.fs

	Pool Name	Size	OSD #	%Data	Target PGs per OSD	Suggested PG Count
■	.rgw.root	3	72	0.10	100	64
■	default.rgw.control	3	72	0.10	100	64
■	default.rgw.data.root	3	72	0.10	100	64
■	default.rgw.gc	3	72	0.10	100	64
■	default.rgw.log	3	72	0.10	100	64
■	default.rgw.intent-log	3	72	0.10	100	64
■	default.rgw.meta	3	72	0.10	100	64
■	default.rgw.usage	3	72	0.10	100	64
■	default.rgw.users.keys	3	72	0.10	100	64
■	default.rgw.users.email	3	72	0.10	100	64
■	default.rgw.users.swift	3	72	0.10	100	64
■	default.rgw.users.uid	3	72	0.10	100	64
■	default.rgw.buckets.extra	3	72	1.00	100	64
■	default.rgw.buckets.index	3	72	3.00	100	64
■	default.rgw.buckets.data	3	72	19.00	100	512
■	kube.rbd	3	72	55.00	100	1024
■	kube.fs	3	72	15.00	100	512
■	kube.dev.rbd	3	72	3.80	100	128
■	kube.dev.fs	3	72	2.00	100	64
				Total Data Percentage: 100.00%		PG Total Count: 3136

相关命令

snippet.bash

```
# ceph osd pool get <poolname> pg_num
# ceph osd pool get <poolname> pgp_num
# ceph osd pool set <poolname> pg_num
# ceph osd pool set <poolname> pgp_num
# ceph osd pool rename <pool> <newpool>
```

```
# ceph pg dump pgs |grep ^<pool number> |awk '{print $1,$2,$17}'
```

使用流程

ceph准备pool和用户

snippet.bash

```
ceph auth get-or-create client.k8s mon 'allow r' osd 'allow rwx pool=k8s,allow rwx pool=rbd' -o  
ceph.client.k8s.keyring
```

创建 secret

snippet.bash

```
# grep key ceph.client.k8s.keyring | awk '{printf "%s", $NF}' | base64  
VBGFaeN3OWJYdUZPSHhBQTNrU2E2QIUyaEF5UUV0SnNPRHdXeRT8PQ==
```

创建类型为 kubernetes.io/rbd 的 Secret

snippet.yaml

```
apiVersion: v1  
kind: Secret  
metadata:  
  name: ceph-k8s-secret  
  type: "kubernetes.io/rbd"  
data:  
  key: VBGFaeN3OWJYdUZPSHhBQTNrU2E2QIUyaEF5UUV0SnNPRHdXeRT8PQ==  
---  
apiVersion: v1  
kind: Secret  
metadata:  
  name: ceph-admin-secret  
  namespace: kube-system  
  type: "kubernetes.io/rbd"  
data:  
  key: VBGFaeN3OWJYdUZPSHhBQTNrU2E2QIUyaEF5UUV0SnNPRHdXeRT8PQ==
```

创建 **StorageClass**

snippet.yaml

```
kind: StorageClass
# 允许扩容
allowVolumeExpansion: true
reclaimPolicy: Retain
metadata:
  name: rbd
provisioner: kubernetes.io/rbd
parameters:
  monitors: 10.32.24.11:6789,10.32.24.12:6789,10.32.24.13:6789
  adminId: k8s
  adminSecretName: ceph-k8s-secret
  adminSecretNamespace: kube-system
  pool: k8s
  userId: k8s
  userSecretName: ceph-k8s-secret
```

创建 **PVC**

snippet.yaml

```
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: test-pvc
spec:
  accessModes:
    - ReadWriteOnce
  storageClassName: rbd
  resources:
    requests:
      storage: 10Gi
```

引用 **PVC**

snippet.yaml

```
apiVersion: apps/v1
kind: Deployment
metadata:
  labels:
```

```
app: ceph-demo
name: ceph-demo
namespace: dev
spec:
  replicas: 1
  template:
    spec:
      containers:
      - env:
        - name: ONLINE_DATE
          value: 2019-09-02
        image: benchmark:latest
        name: nginx
      volumeMounts:
      - mountPath: "/data"
        name: test
      volumes:
      - name: test
        persistentVolumeClaim:
          claimName: test-pvc
```

生命周期管理

操作	效果	备注
删除未挂载的pvc	回收策略未Delete时，pv和ceph image都被删除	比较危险，应人工执行
删除挂载中的pvc	回收策略未Delete时，Pod销毁后pv和ceph image都被删除	比较危险，应人工执行

更改pv回收策略

参考：<https://kubernetes.io/zh/docs/tasks/administer-cluster/change-pv-reclaim-policy/>

snippet.bash

```
kubectl patch pv <your-pv-name> -p '{"spec":{"persistentVolumeReclaimPolicy":"Retain"}}'
```

更改storageclass回收策略

无法patch，似乎可以直接删除在重建

问题

pvc扩容报错

<https://github.com/kubernetes/kubernetes/issues/72393>

升级 kube-controller-manager 到1.12.10版已解决（未完全解决，见
<https://github.com/kubesphere/kubesphere/issues/1156#issuecomment-546223518>）

auth: unable to find a keyring

不需要在 k8s node 上部署 /etc/ceph/ceph.conf 和 /etc/ceph/client.<user>.keyring。报错是因为 secret 中存储的 key 有错误。参考：<https://github.com/kubernetes/kubernetes/issues/66757>

```
auth: unable to find a keyring on /etc/ceph/ceph.client.k8s-  
production.keyring,/etc/ceph/ceph.keyring,/etc/ceph/keyring,/etc/ceph/keyring.bin,: (2) No such file  
or directory
```

检查以下项目

1. base64的字符串不能有换行符，`echo -n <keyring> |base64`
2. 检查用户和pool是否匹配，是否有访问pool的权限

Printed on: 2022/10/05 18:11

Convert to img Failed!