Red Hat OpenShift Container Storage 4.8

恢复 Metro-DR 扩展集群

集群和存储管理员的灾难恢复任务

Last Updated: 2023-02-20

Red Hat OpenShift Container Storage 4.8 恢复 Metro-DR 扩展集群

集群和存储管理员的灾难恢复任务

Enter your first name here. Enter your surname here.
Enter your organisation's name here. Enter your organisational division here.
Enter your email address here.

法律通告

摘要

本文档介绍了如何在 Red Hat OpenShift Container Storage 出现区域灾难时进行恢复。 This is a technology preview feature and is available only for deployments using local storage devices. Technology Preview features are not supported with Red Hat production service level agreements (SLAs) and might not be functionally complete. Red Hat does not recommend using them in production. These features provide early access to upcoming product features, enabling customers to test functionality and provide feedback during the development process.

目录

使开源包含更多

红帽承诺替换我们的代码、文档和网页属性中存在问题的语言。我们从这四个术语开始： master、slave、blacklist 和 whitelist。这些更改将在即将发行的几个发行本中逐渐实施。如需了解更多详细信息，请参阅 CTO Chris Wright 信息。

对红帽文档提供反馈

我们感谢您对文档提供反馈信息。请告诉我们如何让它更好。提供反馈：

- 关于特定内容的简单评论：
  1. 请确定您使用 *Multi-page HTML* 格式查看文档。另外，确定 **Feedback** 按钮出现在文档页的右上方。
  2. 用鼠标指针高亮显示您想评论的文本部分。
  3. 点在高亮文本上弹出的 **Add Feedback**。
  4. 按照显示的步骤操作。
- 要提交更复杂的反馈，请创建一个 Bugzilla ticket：
  1. 进入 Bugzilla 网站。
  2. 在 Component 中选择 **Documentation**。
  3. 在 **Description** 中输入您要提供的信息。包括文档相关部分的链接。
  4. 点 **Submit Bug**。

第 1 章 概述

鉴于扩展集群的区域灾难恢复功能是在环境出现完全或部分故障时为环境提供抗压性，因此了解应用程序及其存储的不同恢复方法非常重要。

应用程序的架构将确定其在活动区域可重新可用的时间。

根据站点的中断情况，应用程序及其存储的恢复方法可能有所不同。恢复时间取决于应用程序架构。不同的恢复方法如下：

- 恢复使用 RWX 存储的区域感知 HA 应用程序
- 使用 RWX 存储的 HA 应用程序的恢复
- 使用 RWO 存储的应用程序恢复
- StatefulSet pod 的恢复

第 2 章 了解区失败

就本节而言，在出现以下情况下我们将把一个区视为故障：一个区中的所有 OpenShift Container Platform 节点、master 节点和 worker 都不再能与第二个数据区中的资源进行通信（例如，节点被关闭）。如果数据区域之间的通信仍处于部分工作状态（不定时地出现关闭/启动情况），集群、存储和网络管理员应采取措施中断用于控制数据区域之间的通信，以确保可以成功进行恢复。

第 3 章 恢复使用 RWX 存储的区域感知 HA 应用程序

使用 **topologyKey: topology.kubernetes.io/zone** 部署、在每个数据区域中有一个或多个副本并且使用共享存储的应用（即 RWX cephfs 卷），可在 30-60 秒内恢复用于新连接。当路由器 Pod 在失败的数据区中离线时，短暂暂停用于 **HAProxy** 刷新连接。

此类型的应用示例在 Install Zone Aware Sample Application 部分中进行了详细说明。

注意

在安装 Sample Application 时，通过关闭 OpenShift Container Platform 节点（至少是使用 OpenShift Container Storage 设备的节点）来验证 file-uploader 应用程序仍可用且可以继续上传新文件，来测试数据区的故障恢复功能。

第 4 章 使用 RWX 存储的 HA 应用程序的恢复

使用 **topologyKey: kubernetes.io/hostname** 或没有拓扑配置的应用程序无法防止同一区域中的所有应用副本。

注意

即使 **Pod** spec 中有 *podAntiAffinity* 和 **topologyKey: kubernetes.io/hostname** 也会发生，因为此反关联性规则基于主机且不是基于区域的规则。

如果发生这种情况，且所有副本都位于失败的区域中，则使用 RWX 存储的应用将需要 6-8 分钟在活动区域中恢复。此暂停时间用于故障区中的 OpenShift Container Platform 节点变为 **NotReady** （60 秒），然后让默认 pod 驱除超时过期（300 秒）。

第 5 章 使用 RWO 存储恢复应用程序

使用 RWO 存储(ReadWriteOnce)的应用程序具有在此 Kubernetes 问题中描述的已知行为。由于此问题，如果数据区出现任何挂载 RWO 卷的应用程序 pod 失败（例如： 基于 **cephrbd** 的卷）在 6-8 分钟后处于 **Terminating** 状态，且不会在没有手动干预的情况下在活动区重新创建。

检查 OpenShift Container Platform 节点，其状态为 **NotReady**。它可能会遇到阻止他们与 OpenShift 控制平面通信的问题。尽管存在这个通信问题，仍然可能对持久卷执行 IO 操作。

如果两个 pod 同时写入同一个 RWO 卷，则存在数据崩溃的风险。必须采取一些措施来确保 **NotReady** 节点上的进程被终止或阻止，直到进程被终止为止。

- 例如，使用具有确认功能的带外管理系统关闭一个节点，是一个确保进程被中断的方法示例。
- 另外一个方法是，撤回故障端点中被节点用于与存储进行通讯的网络路由器。

注意

在将服务恢复到失败的区域或节点前，必须先确认所有带有持久性卷的 pod 都已成功终止。

要让 **Terminating** pod 在活跃区中重新创建，您可以强制删除 pod 或删除关联的 PV 上的终结器。完成这两个操作之一后，应用程序 Pod 应在 active 区域中重新创建，并成功挂载其 RWO 存储。

强制删除 pod

强制删除不需要等待 kubelet 确认 Pod 已终止的信息。

```
$ oc delete pod <PODNAME> --grace-period=0 --force --namespace <NAMESPACE>
```

<PODNAME>

是 pod 的名称

<NAMESPACE>

是项目的命名空间

删除关联的 PV 上的终结器

找到由 **Terminating** pod 挂载的持久性卷声明(PVC)关联的 PV，并使用 **oc patch** 命令删除终结器。

```
$ oc patch -n openshift-storage pv/<PV_NAME> -p '{"metadata":{"finalizers":[]}}' --type=merge
```

<PV_NAME>

是 PV 的名称

查找关联的 PV 的一种简单方法是描述 Terminating pod。如果您看到多附件警告，在警告中应具有 PV 名称（例如： pvc-0595a8d2-683f-443b-aee0-6e547f5f5a7c）。

```
$ oc describe pod <PODNAME> --namespace <NAMESPACE>
```

<PODNAME>

是 pod 的名称

<NAMESPACE>

是项目的命名空间

输出示例：

```
[...]
Events:
  Type     Reason                  Age   From                     Message
  ----     ------                  ----  ----                     -------
  Normal   Scheduled               4m5s  default-scheduler        Successfully assigned openshift-
storage/noobaa-db-pg-0 to perf1-mz8bt-worker-d2hdm
  Warning  FailedAttachVolume      4m5s  attachdetach-controller  Multi-Attach error for volume
"pvc-0595a8d2-683f-443b-aee0-6e547f5f5a7c" Volume is already exclusively attached to one
node and can't be attached to another
```

第 6 章 STATEFULSET POD 的恢复

属于有状态集合的 Pod 与 Pod 挂载 RWO 卷有类似的问题。如需更多信息，请参阅 Kubernetes 资源 StatefulSet 注意事项。

要让 StatefulSet 的 Pod 部分在 6-8 分钟后重新创建在活跃区中，需要强制删除 Pod，要求与具有 RWO 卷的 Pod 相同要求（例如，OpenShift Container Platform 节点关机或中断通信）。