



Red Hat Ceph Storage 7

OpenStack 块设备指南

配置 Ceph、QEMU、libvirt 和 OpenStack 以使用 Ceph 作为 OpenStack 的后端。

Red Hat Ceph Storage 7 OpenStack 块设备指南

配置 Ceph、QEMU、libvirt 和 OpenStack 以使用 Ceph 作为 OpenStack 的后端。

法律通告

Copyright © 2024 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux[®] is the registered trademark of Linus Torvalds in the United States and other countries.

Java[®] is a registered trademark of Oracle and/or its affiliates.

XFS[®] is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL[®] is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js[®] is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack[®] Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

摘要

本文档论述了如何将 OpenStack 和 Ceph 配置为使用 Ceph 作为 Glance、Cinder、Cinder 备份和 Nova 的后端。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、黑名单和白名单。由于此项工作十分艰巨，这些更改将在即将推出的几个发行版本中逐步实施。详情请查看 CTO Chris Wright 信息

目录

第1章 CEPH 块设备和 OPENSTACK	3
第2章 为 OPENSTACK 安装和配置 CEPH	5
2.1. 为 OPENSTACK 创建 CEPH 池	5
2.2. 在 OPENSTACK 上安装 CEPH 客户端	5
2.3. 将 CEPH 配置文件复制到 OPENSTACK	6
2.4. 配置 CEPH 客户端身份验证	6
第3章 将 OPENSTACK 配置为使用 CEPH 块设备	9
3.1. 将 CINDER 配置为使用 CEPH 块设备	9
3.2. 配置 CINDER 备份以使用 CEPH 块设备	10
3.3. 将 GLANCE 配置为使用 CEPH 块设备	12
3.4. 将 NOVA 配置为使用 CEPH 块设备	13
3.5. 重启 OPENSTACK 服务	14

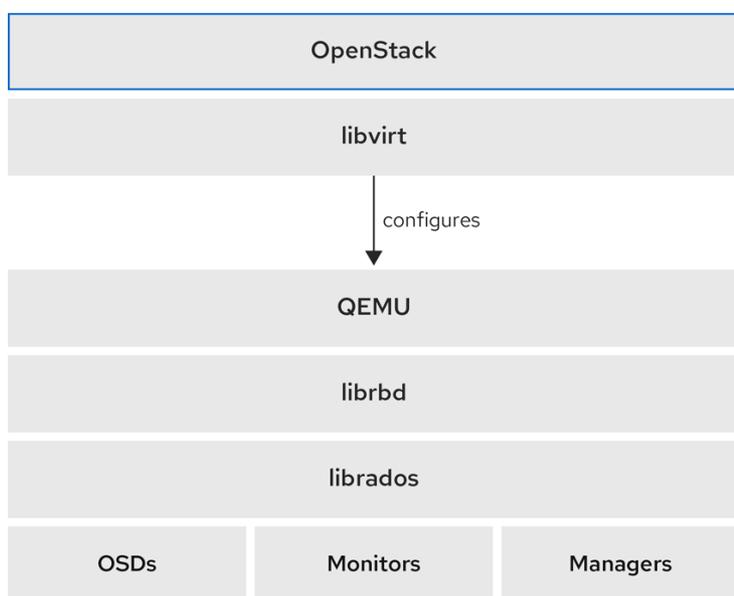
第1章 CEPH 块设备和 OPENSTACK

Red Hat Enterprise Linux OpenStack Platform Director 提供了两种方法，将 Ceph 用作 Glance 的后端，即 Cinder、Cinder 备份和 Nova：

1. **OpenStack 创建 Ceph 存储集群**：OpenStack Director 可以创建 Ceph 存储集群。这需要为 Ceph OSD 配置模板。OpenStack 处理 Ceph 主机的安装和配置。借助这种情形，OpenStack 将使用 OpenStack 控制器主机安装 Ceph 监视器。
2. **OpenStack 连接到现有的 Ceph 存储集群**：OpenStack Director 使用 Red Hat OpenStack Platform 9 及更高版本，可以连接到 Ceph 监视器并配置 Ceph 存储集群，以用作 OpenStack 的后端。

以上方法是将 Ceph 配置为 OpenStack 后端的首选方法，因为它们将自动处理大部分的安装和配置。

本文档详细介绍了配置 Ceph、QEMU、libvirt 和 OpenStack 的手动流程，以将 Ceph 用作后端。本文档适用于不打算使用 RHEL OSP Director 的用户。



112_Ceph_0720



注意

正在运行的 Ceph 存储集群和至少一个 OpenStack 主机需要使用 Ceph 块设备作为 OpenStack 的后端。

OpenStack 的三个部分与 Ceph 的块设备集成：

- **镜像**：OpenStack Glance 管理虚拟机的镜像。镜像是不可变的，OpenStack 将镜像视为二进制 blob 并相应地进行下载。
- **卷**：卷是块设备。OpenStack 使用卷来引导虚拟机，或把卷附加到运行的虚拟机上。OpenStack 使用 Cinder 服务管理卷。Ceph 可以充当 OpenStack Cinder 和 Cinder 备份的黑色结尾。
- **客户机磁盘**：客户机磁盘是客户机操作系统磁盘。默认情况下，在引导虚拟机时，默认情况下，其磁盘会在虚拟机监控程序的文件系统上显示为一个文件，默认在 `/var/lib/nova/instances/<uuid>/` 目录中。OpenStack Glance 可以将镜像存储在 Ceph 块设备中，并且可以使用 Cinder 使用镜像的写时复制克隆来引导虚拟机。



重要

Ceph 不支持 QCOW2 托管虚拟机磁盘。要引导虚拟机，临时后端或从卷引导，Glance 镜像格式必须是 RAW。

OpenStack 可以使用 Ceph 作为镜像、卷或客户机磁盘虚拟机。但并不需要全部 3 个都使用。

其它资源

- 如需了解更多详细信息，请参阅 [Red Hat OpenStack Platform 文档](#)。

第 2 章 为 OPENSTACK 安装和配置 CEPH

作为存储管理员，您必须在 Red Hat OpenStack Platform 使用 Ceph 块设备前安装和配置 Ceph。

先决条件

- 新的或现有的 Red Hat Ceph Storage 集群。

2.1. 为 OPENSTACK 创建 CEPH 池

您可以创建 Ceph 池以用于 OpenStack。默认情况下，Ceph 块设备使用 **rbd** 池，但您可以使用任何可用的池。

先决条件

- 一个正在运行的 Red Hat Ceph Storage 集群。

流程

1. 验证 Red Hat Ceph Storage 集群是否正在运行，且处于 **HEALTH_OK** 状态：

```
[root@mon ~]# ceph -s
```

2. 创建 Ceph 池：

示例

```
[root@mon ~]# ceph osd pool create volumes 128
[root@mon ~]# ceph osd pool create backups 128
[root@mon ~]# ceph osd pool create images 128
[root@mon ~]# ceph osd pool create vms 128
```

在上例中，**128** 是放置组的数量。



重要

红帽建议使用 [Ceph Placement Group's per Pool Calculator](#) 来计算池的适当放置组数量。

其它资源

- 有关创建 [池](#) 的详情，请参阅 *存储策略指南* 中的池章节。

2.2. 在 OPENSTACK 上安装 CEPH 客户端

您可以在 Red Hat OpenStack Platform 上安装 Ceph 客户端软件包，以访问 Ceph 存储集群。

先决条件

- 一个正在运行的 Red Hat Ceph Storage 集群。
- 访问 Ceph 软件存储库。

- OpenStack Nova、Cinder、Cinder 备份和 Glance 节点的 root 级别访问权限。

流程

1. 在 OpenStack Nova 上，Cinder 备份节点安装以下软件包：

```
[root@nova ~]# dnf install python-rbd ceph-common
```

2. 在 OpenStack Glance 主机上安装 **python-rbd** 软件包：

```
[root@glance ~]# dnf install python-rbd
```

2.3. 将 CEPH 配置文件复制到 OPENSTACK

将 Ceph 配置文件复制到 **nova-compute**、**cinder-backup**、**cinder-volume** 和 **glance-api** 节点。

先决条件

- 一个正在运行的 Red Hat Ceph Storage 集群。
- 访问 Ceph 软件存储库。
- OpenStack Nova、Cinder 和 Glance 节点的 root 级别访问权限。

流程

1. 将 Ceph 配置文件从 Ceph 监控主机复制到 OpenStack Nova、Cinder、Cinder 备份和 Glance 节点：

```
[root@mon ~]# scp /etc/ceph/ceph.conf OPENSTACK_NODES:/etc/ceph
```

2.4. 配置 CEPH 客户端身份验证

您可以为 Ceph 客户端配置身份验证以访问 Red Hat OpenStack Platform。

先决条件

- Ceph 监控主机的 root 级别访问权限。
- 一个正在运行的 Red Hat Ceph Storage 集群。

流程

1. 从 Ceph 监控主机，为 Cinder、Cinder 备份和 Glance 创建新用户：

```
[root@mon ~]# ceph auth get-or-create client.cinder mon 'allow r' osd 'allow class-read
object_prefix rbd_children, allow rwx pool=volumes, allow rwx pool=vms, allow rx
pool=images'
```

```
[root@mon ~]# ceph auth get-or-create client.cinder-backup mon 'allow r' osd 'allow class-
read object_prefix rbd_children, allow rwx pool=backups'
```

```
[root@mon ~]# ceph auth get-or-create client.glance mon 'allow r' osd 'allow class-read
object_prefix rbd_children, allow rwx pool=images'
```

2. 将 **client.cinder**、**client.cinder-backup** 和 **client.glance** 的密钥环添加到适当的节点，并更改其所有权：

```
[root@mon ~]# ceph auth get-or-create client.cinder | ssh CINDER_VOLUME_NODE sudo
tee /etc/ceph/ceph.client.cinder.keyring
[root@mon ~]# ssh CINDER_VOLUME_NODE chown cinder:cinder
/etc/ceph/ceph.client.cinder.keyring
```

```
[root@mon ~]# ceph auth get-or-create client.cinder-backup | ssh CINDER_BACKUP_NODE
tee /etc/ceph/ceph.client.cinder-backup.keyring
[root@mon ~]# ssh CINDER_BACKUP_NODE chown cinder:cinder
/etc/ceph/ceph.client.cinder-backup.keyring
```

```
[root@mon ~]# ceph auth get-or-create client.glance | ssh GLANCE_API_NODE sudo tee
/etc/ceph/ceph.client.glance.keyring
[root@mon ~]# ssh GLANCE_API_NODE chown glance:glance
/etc/ceph/ceph.client.glance.keyring
```

3. OpenStack Nova 节点需要 **nova-compute** 进程的密钥环文件：

```
[root@mon ~]# ceph auth get-or-create client.cinder | ssh NOVA_NODE tee
/etc/ceph/ceph.client.cinder.keyring
```

4. OpenStack Nova 节点还需要将 **client.cinder** 用户的 secret 密钥存储在 **libvirt** 中。**libvirt** 进程需要 secret 密钥来访问集群，同时从 Cinder 附加块设备。在 OpenStack Nova 节点上创建 secret 密钥的临时副本：

```
[root@mon ~]# ceph auth get-key client.cinder | ssh NOVA_NODE tee client.cinder.key
```

如果存储集群包含使用 **exclusive-lock** 功能的 Ceph 块设备镜像，请确保所有 Ceph 块设备用户都有 blocklist 客户端的权限：

```
[root@mon ~]# ceph auth caps client.ID mon 'allow r, allow command "osd blacklist"' osd
'EXISTING_OSD_USER_CAPS'
```

5. 返回到 OpenStack Nova 主机：

```
[root@mon ~]# ssh NOVA_NODE
```

6. 为 secret 生成 UUID，并保存用于配置 **nova-compute** 的 secret 的 UUID：

```
[root@nova ~]# uuidgen > uuid-secret.txt
```



注意

您不一定需要所有 Nova 计算节点上的 UUID。但是，从平台一致性角度来说，最好保留相同的 UUID。

- 在 OpenStack Nova 节点上，将 secret 密钥添加到 **libvirt** 中，并删除密钥的临时副本：

```
cat > secret.xml <<EOF
<secret ephemeral='no' private='no'>
  <uuid>`cat uuid-secret.txt`</uuid>
  <usage type='ceph'>
    <name>client.cinder secret</name>
  </usage>
</secret>
EOF
```

- 为 **libvirt** 设置并定义 secret：

```
[root@nova ~]# virsh secret-define --file secret.xml
[root@nova ~]# virsh secret-set-value --secret $(cat uuid-secret.txt) --base64 $(cat
client.cinder.key) && rm client.cinder.key secret.xml
```

其它资源

- 如需了解更多详细信息，请参阅 *Red Hat Ceph Storage Administration Guide* 中的 [Managing Ceph users](#) 章节。
- 以了解更多有关用户功能的信息，请参阅 *Integrating an Overcloud with an Existing Red Hat Ceph Cluster Guide* 中的 [配置现有 ceph 存储集群](#) 一节。

第 3 章 将 OPENSTACK 配置为使用 CEPH 块设备

作为存储管理员，您必须将 Red Hat OpenStack Platform 配置为使用 Ceph 块设备。Red Hat OpenStack Platform 可以使用 Ceph 块设备用于 Cinder、Cinder 备份、Glance 和 Nova。

先决条件

- 新的或现有的 Red Hat Ceph Storage 集群。
- 正在运行的 Red Hat OpenStack Platform 环境。

3.1. 将 CINDER 配置为使用 CEPH 块设备

Red Hat OpenStack Platform 可以使用 Ceph 块设备为 Cinder 卷提供后端存储。

先决条件

- Cinder 节点的 root 级别访问权限。
- Ceph 卷池。
- 与 Ceph 块设备交互的机密的用户和 UUID。

流程

1. 编辑 Cinder 配置文件：

```
[root@cinder ~]# vim /etc/cinder/cinder.conf
```

2. 在 **[DEFAULT]** 部分中，将 Ceph 启用为 Cinder 的后端：

```
enabled_backends = ceph
```

3. 确保 Glance API 版本设为 2。如果要在 **enabled_backends** 中配置多个 cinder 后端，**glance_api_version = 2** 设置必须位于 **[DEFAULT]** 部分中，而不是 **[ceph]** 部分。

```
glance_api_version = 2
```

4. 在 **cinder.conf** 文件中创建一个 **[ceph]** 部分。在以下步骤的 **[ceph]** 部分下，添加 Ceph 设置。

5. 指定 **volume_driver** 设置，并将其设置为使用 Ceph 块设备驱动程序：

```
volume_driver = cinder.volume.drivers.rbd.RBDDriver
```

6. 指定集群名称和 Ceph 配置文件的位置。在典型的部署中，Ceph 集群具有集群名称 **ceph** 和 Ceph 配置文件（位于 **/etc/ceph/ceph.conf**）。如果 Ceph 集群名称不是 **ceph**，请相应地指定集群名称和配置文件路径：

```
rbd_cluster_name = us-west
rbd_ceph_conf = /etc/ceph/us-west.conf
```

7. 默认情况下，Red Hat OpenStack Platform 将 Ceph 卷存储在 **rbd** 池中。要使用之前创建的 **volumes** 池，请指定 **rbd_pool** 设置并设置 **volumes** 池：

```
rbd_pool = volumes
```

- Red Hat OpenStack Platform 没有卷的默认用户名或 secret 的 UUID。指定 **rbd_user**，并将它设置为 **cinder** 用户。然后，指定 **rbd_secret_uuid** 设置，并将其设置为 **uuid-secret.txt** 文件中存储的生成的 UUID：

```
rbd_user = cinder
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
```

- 指定以下设置：

```
rbd_flatten_volume_from_snapshot = false
rbd_max_clone_depth = 5
rbd_store_chunk_size = 4
rados_connect_timeout = -1
```

当您将在 Cinder 配置为使用 Ceph 块设备时，配置文件可能类似如下：

示例

```
[DEFAULT]
enabled_backends = ceph
glance_api_version = 2
...

[ceph]
volume_driver = cinder.volume.drivers.rbd.RBDDriver
rbd_cluster_name = ceph
rbd_pool = volumes
rbd_user = cinder
rbd_ceph_conf = /etc/ceph/ceph.conf
rbd_flatten_volume_from_snapshot = false
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
rbd_max_clone_depth = 5
rbd_store_chunk_size = 4
rados_connect_timeout = -1
```



注意

考虑删除默认的 **[lvm]** 部分及其设置。

3.2. 配置 CINDER 备份以使用 CEPH 块设备

Red Hat OpenStack Platform 可以配置 Cinder 备份以使用 Ceph 块设备。

先决条件

- Cinder 节点的 root 级别访问权限。

流程

- 编辑 Cinder 配置文件：

```
[root@cinder ~]# vim /etc/cinder/cinder.conf
```

2. 前往配置文件的 **[ceph]** 部分。
3. 指定 **backup_driver** 设置，并将其设置为 Ceph 驱动程序：

```
backup_driver = cinder.backup.drivers.ceph
```

4. 指定 **backup_ceph_conf** 设置并指定 Ceph 配置文件的路径：

```
backup_ceph_conf = /etc/ceph/ceph.conf
```

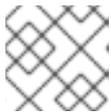


注意

Cinder 备份 Ceph 配置文件可能与用于 Cinder 的 Ceph 配置文件不同。例如，它可以指向不同的 Ceph 存储集群。

5. 指定用于备份的 Ceph 池：

```
backup_ceph_pool = backups
```



注意

用于 Cinder 备份的 Ceph 配置文件可能与用于 Cinder 的 Ceph 配置文件不同。

6. 指定 **backup_ceph_user** 设置，并将用户指定为 **cinder-backup**：

```
backup_ceph_user = cinder-backup
```

7. 指定以下设置：

```
backup_ceph_chunk_size = 134217728
backup_ceph_stripe_unit = 0
backup_ceph_stripe_count = 0
restore_discard_excess_bytes = true
```

当您包含 Cinder 选项时，**cinder.conf** 文件的 **[ceph]** 部分可能类似如下：

示例

```
[ceph]
volume_driver = cinder.volume.drivers.rbd.RBDDriver
rbd_cluster_name = ceph
rbd_pool = volumes
rbd_user = cinder
rbd_ceph_conf = /etc/ceph/ceph.conf
rbd_flatten_volume_from_snapshot = false
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
rbd_max_clone_depth = 5
rbd_store_chunk_size = 4
rados_connect_timeout = -1
```

```

backup_driver = cinder.backup.drivers.ceph
backup_ceph_user = cinder-backup
backup_ceph_conf = /etc/ceph/ceph.conf
backup_ceph_chunk_size = 134217728
backup_ceph_pool = backups
backup_ceph_stripe_unit = 0
backup_ceph_stripe_count = 0
restore_discard_excess_bytes = true

```

8. 验证是否启用了 Cinder 备份：

```
[root@cinder ~]# cat /etc/openstack-dashboard/local_settings | grep enable_backup
```

如果 **enable_backup** 设为 **False**，则编辑 **local_settings** 文件并将其设置为 **True**。

示例

```

OPENSTACK_CINDER_FEATURES = {
    'enable_backup': True,
}

```

3.3. 将 GLANCE 配置为使用 CEPH 块设备

Red Hat OpenStack Platform 可以将 Glance 配置为使用 Ceph 块设备。

先决条件

- 对 Glance 节点的 root 级别访问权限。

流程

1. 要默认使用 Ceph 块设备，请编辑 **/etc/glance/glance-api.conf** 文件。如果您使用了不同的池，用户或 Ceph 配置文件设置会应用适当的值。如有必要，取消注释以下设置，并相应地更改其值：

```
[root@glance ~]# vim /etc/glance/glance-api.conf
```

```

stores = rbd
default_store = rbd
rbd_store_chunk_size = 8
rbd_store_pool = images
rbd_store_user = glance
rbd_store_ceph_conf = /etc/ceph/ceph.conf

```

2. 要启用 copy-on-write (CoW)克隆，请将 **show_image_direct_url** 设置为 **True**。

```
show_image_direct_url = True
```



重要

启用 CoW 通过 Glance 的 API 公开后端位置，因此不应公开访问端点。

3. 如果需要，禁用缓存管理。类别 应仅设置为 **keystone**，而不是 **keystone+cachemanagement**。

```
flavor = keystone
```

4. 红帽建议镜像的以下属性：

```
hw_scsi_model=virtio-scsi
hw_disk_bus=scsi
hw_qemu_guest_agent=yes
os_require_quiesce=yes
```

virtio-scsi 控制器获得更好的性能，并提供对丢弃操作的支持。对于使用 SCSI/SAS 驱动器的系统，请将每个 Cinder 块设备连接到该控制器。另外，启用 QEMU 客户机代理并通过 QEMU 客户机代理发送 **fs-freeze/thaw** 调用。

3.4. 将 NOVA 配置为使用 CEPH 块设备

Red Hat OpenStack Platform 可以将 Nova 配置为使用 Ceph 块设备。

您必须将每个 Nova 节点配置为使用临时后端存储设备，允许所有虚拟机使用 Ceph 块设备。

先决条件

- 对 Nova 节点的 root 级别访问权限。

流程

1. 编辑 Ceph 配置文件：

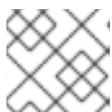
```
[root@nova ~]# vim /etc/ceph/ceph.conf
```

2. 将以下部分添加到 Ceph 配置文件的 **[client]** 部分：

```
[client]
rbd cache = true
rbd cache writethrough until flush = true
rbd concurrent management ops = 20
admin socket = /var/run/ceph/guests/$cluster-$type.$id.$pid.$cctid.asok
log file = /var/log/ceph/qemu-guest-$pid.log
```

3. 为 admin 套接字和日志文件创建新目录，并将目录权限更改为使用 **qemu** 用户和 **libvirt** 组：

```
[root@nova ~]# mkdir -p /var/run/ceph/guests/ /var/log/ceph/
[root@nova ~]# chown qemu:libvirt /var/run/ceph/guests /var/log/ceph/
```



注意

SELinux 或 AppArmor 必须允许该目录。

4. 在每个 Nova 节点上，编辑 **/etc/nova/nova.conf** 文件。在 **[libvirt]** 部分下配置以下设置：

示例

```
[libvirt]
images_type = rbd
images_rbd_pool = vms
images_rbd_ceph_conf = /etc/ceph/ceph.conf
rbd_user = cinder
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
disk_cachemodes="network=writeback"
inject_password = false
inject_key = false
inject_partition = -2
live_migration_flag="VIR_MIGRATE_UNDEFINE_SOURCE,VIR_MIGRATE_PEER2PEER,VIR_MIGRATE_LIVE,VIR_MIGRATE_PERSIST_DEST,VIR_MIGRATE_TUNNELLED"
hw_disk_discard = unmap
```

将 **rbd_user_secret** 中的 UUID 替换为 **uuid-secret.txt** 文件中的 UUID。

3.5. 重启 OPENSTACK 服务

重启 Red Hat OpenStack Platform 服务可让您激活 Ceph 块设备驱动程序。

先决条件

- 对 Red Hat OpenStack Platform 节点的 root 级别访问权限。

流程

1. 将块设备池名称和 Ceph 用户名加载到配置文件中。
2. 在修改对应的配置文件后重启适当的 OpenStack 服务：

```
[root@osp ~]# systemctl restart openstack-cinder-volume
[root@osp ~]# systemctl restart openstack-cinder-backup
[root@osp ~]# systemctl restart openstack-glance-api
[root@osp ~]# systemctl restart openstack-nova-compute
```