Abstract

This document contains information about the Ceph technology preview release available from Oracle. It describes the differences from the upstream version, includes notes on installing and configuring Ceph, and provides a statement of what is supported.

Document generated on: 2015-10-16 (revision: 3347)
# Table of Contents

Preface ....................................................................................................................................................... v  
1 Release Notes ......................................................................................................................................... 1  
  1.1 About Ceph for Oracle Linux .................................................................................................... 1  
  1.2 About Ceph .................................................................................................................................... 1  
  1.3 Differences from the Upstream Release .................................................................................... 1  
  1.4 Enabling Access to the Ceph Packages on Oracle Public Yum ............................................ 1  
  1.5 Preparing the Storage Cluster Nodes Before Installing Ceph ................................................ 2  
  1.6 Installing and Configuring Ceph on the Storage Cluster Deployment Node ............................ 3  
  1.7 Installing Ceph on the Other Storage Cluster Nodes .............................................................. 4  
  1.8 Configuring the Storage Cluster ............................................................................................... 4  
  1.9 Configuring an Object Gateway for OpenStack and Swift Access ....................................... 6  
  1.10 Installing a Ceph Client ............................................................................................................ 6  
  1.11 Creating a Storage Pool ............................................................................................................ 7  
  1.12 Configuring a Block Device on a Ceph Client ......................................................................... 7  
  1.13 Known Issues .............................................................................................................................. 7  
    1.13.1 ceph-deploy Reports Errors ................................................................................................. 8  
    1.13.2 ceph-deploy Reports an Error on Exit .............................................................................. 8  
    1.13.3 ceph-deploy Python Errors on Oracle Linux 6 .............................................................. 8  
    1.13.4 ceph-radgosgw Service Fails to Start on Oracle Linux 7 ............................................ 8  
    1.13.5 ceph-radgosgw Service Produces Error and Warning Messages when Starting on Oracle Linux 6 .......................................................... 8  
    1.13.6 Mounting a CephFS File System Produces a Spurious Error Message ....................... 9  
    1.13.7 sosreport Takes a Long Time to Run on a Storage Cluster Node .................................. 9  
Ceph Terminology .................................................................................................................................. 11
Preface

The Ceph for Oracle Linux Release Notes provides details of the Firefly release of Ceph that is available from Oracle as a technology preview for Oracle Linux 6 and Oracle Linux 7.

Audience

This document is written for developers who want to use Ceph with Oracle Linux 6 or Oracle Linux 7. It is assumed that readers have a general understanding of the Linux operating system.

Related Documents

The latest version of this document and other documentation for this product are available at:


Conventions

The following text conventions are used in this document:

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>boldface</td>
<td>Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.</td>
</tr>
<tr>
<td>italic</td>
<td>Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.</td>
</tr>
<tr>
<td>monospace</td>
<td>Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.</td>
</tr>
</tbody>
</table>
Chapter 1 Release Notes

1.1 About Ceph for Oracle Linux

Ceph is currently available as a technology preview for Oracle Linux 6 Update 7 (x86_64) or later and Oracle Linux 7 Update 1 (x86_64) or later with the Unbreakable Enterprise Kernel Release 3 Quarterly Update 6 or later.

Note
The source RPMs for Ceph are available from Oracle Public Yum at http://public-yum.oracle.com/beta.

This technical preview of Ceph for Oracle Linux is based on the Ceph Community Firefly release (v0.80). It includes the Object Store, Block Device, Storage Cluster, and Object Gateway components of Ceph. The Ceph file system (CephFS) component is also included but it is not a technical preview feature.

For a quick-start guide to using Ceph, see http://ceph.com/docs/master/start/quick-ceph-deploy/.

For more information about Ceph, go to http://ceph.com/.

1.2 About Ceph

Ceph presents a uniform view of object and block storage from a cluster of multiple physical and logical commodity-hardware storage devices. Ceph can provide fault tolerance and enhance I/O performance by replicating and striping data across the storage devices in a Storage Cluster. Ceph's monitoring and self-repair features minimize administration overhead. You can configure a Storage Cluster on non-identical hardware from different manufacturers.

1.3 Differences from the Upstream Release

There are no differences between the Oracle versions of the software collections and the upstream release apart from the addition of Oracle Linux GPG keys.

1.4 Enabling Access to the Ceph Packages on Oracle Public Yum

The ceph-deploy package is available on Oracle Public Yum or ULN. This procedure describes how to enable access to Oracle Public Yum for the system that will act as the Storage Cluster deployment node.

Perform the following steps on the system that will act as the Storage Cluster deployment node:

1. Use a command such as curl or wget to download a yum repository file that includes the:
   
   - For example, using wget with Oracle Linux 6:
   ```bash
   # wget -O /etc/yum.repos.d/public-yum-ol6-ceph-beta.repo \
   
   # For example, using wget with Oracle Linux 7:
   ```
   ```bash
   # wget -O /etc/yum.repos.d/public-yum-ol7-ceph-beta.repo \
   ```

   The downloaded file contains enabled entries for the ol6_ceph10_beta or ol7_ceph10_beta repository as appropriate.
2. The downloaded file contains disabled entries for additional repositories that contain packages on which Ceph depends:

- Oracle Linux 6: \texttt{ol6\_latest}
- Oracle Linux 7: \texttt{ol7\_latest} and \texttt{ol7\_optional\_latest}

If the system does not already have access to these repositories on Public Yum or ULN, you can enable access by editing the downloaded yum repository file:

- On Oracle Linux 6, set the values of the \texttt{enabled} parameter for the \texttt{ol6\_latest} repository to 1.
- On Oracle Linux 7, set the values of the \texttt{enabled} parameters for the \texttt{ol7\_latest} and \texttt{ol7\_optional\_latest} repositories to 1.

3. For Oracle Linux 7, enable access to the \texttt{ol7\_addons} repository on Public Yum or ULN.

You can now prepare the Storage Cluster nodes for Ceph installation. See Section 1.5, “Preparing the Storage Cluster Nodes Before Installing Ceph”.

\section*{1.5 Preparing the Storage Cluster Nodes Before Installing Ceph}

\begin{mdframed}
\textbf{Note}

For data integrity, a Storage Cluster should contain two or more nodes for storing copies of an object.

For high availability, a Storage Cluster should contain three or more nodes that store copies of an object.

In the example used in the following steps, the administration node is \texttt{ceph\_node1.mydom.com (192.168.1.51)}.

To prepare the Storage Cluster nodes:

1. Perform the following steps on each system that will act as a Storage Cluster node:

a. If SELinux is enabled, disable it and reboot the system:

\begin{verbatim}
\texttt{$\ downarrow$ sed -i '/SELINUX/s/enforcing/disabled/' /etc/selinux/config}
\end{verbatim}

You can re-enable SELinux on each node after configuring Ceph.

b. If the NTP service is not already configured, install and start it. See the \texttt{Oracle Linux 6 Administrator's Guide} or the \texttt{Oracle Linux 7 Administrator's Guide} as appropriate.

\begin{mdframed}
\textbf{Note}

Use the \texttt{hwclock --show} command to ensure that all nodes agree on the time. By default, the Ceph monitors report \texttt{health HEALTH\_WARN clock skew detected on mon} errors if the clocks on the nodes differ by more than 50 milliseconds.
\end{mdframed}

c. Stop and disable the firewall service.

For Oracle Linux 6, enter:

\begin{verbatim}
\texttt{$\ downarrow$ service iptables stop}
\end{verbatim}
\end{mdframed}
# service ip6tables stop
# chkconfig iptables off
# chkconfig ip6tables off

For Oracle Linux 7, enter:

# systemctl stop firewalld
# systemctl disable firewalld

You can restart, re-enable and reconfigure the firewall on each node after configuring Ceph.

d. Edit /etc/hosts and add entries for the IP address and host name of all of the nodes in the Storage Cluster, for example:

192.168.1.51    ceph-node1.mydom.com  ceph-node1
192.168.1.52    ceph-node2.mydom.com  ceph-node2
192.168.1.53    ceph-node3.mydom.com  ceph-node3
192.168.1.54    ceph-node4.mydom.com  ceph-node4

Note

Although you can use DNS to configure host name to IP address mapping, Oracle recommends that you also configure /etc/hosts in case the DNS service becomes unavailable.

2. Enable SSH on the nodes:

a. On the administration node, generate the SSH key, specifying an empty passphrase:

# ssh-keygen

b. Copy the key to the other nodes in the Storage Cluster, for example:

# ssh-copy-id root@ceph-node2
# ssh-copy-id root@ceph-node3
# ssh-copy-id root@ceph-node4

You can now install and configure the Storage Cluster deployment node, which is usually the same system as the administration node. See Section 1.6, “Installing and Configuring Ceph on the Storage Cluster Deployment Node”.

1.6 Installing and Configuring Ceph on the Storage Cluster Deployment Node

Note

In the example used in the following steps, the deployment node is ceph-node1.mydom.com (192.168.1.51), which is the same as the administration node.

Perform the following steps on the deployment node:

1. Install the ceph-deploy package.

# yum install ceph-deploy

2. Create a Ceph configuration directory for the Storage Cluster and change to this directory, for example:

# mkdir /var/mydom_ceph
# cd /var/mydom_ceph
3. Use the `ceph-deploy` command to define the members of the Storage Cluster, for example:

```
# ceph-deploy --cluster mydom_ceph new ceph-node{1,2,3,4}
```

**Note**
If you do not intend to run more than one Storage Cluster on the same hardware, you do not need to specify a cluster name using the `--cluster` option.

4. Edit `/etc/ceph/ceph.conf` and set the default number of replicas, for example:

```
osd pool default size = 3
```

You can now install Ceph on the remaining Storage Cluster nodes. See Section 1.7, “Installing Ceph on the Other Storage Cluster Nodes”.

### 1.7 Installing Ceph on the Other Storage Cluster Nodes

Having installed and configured the Ceph deployment node, you can use this node to install Ceph on the other nodes.

To install Ceph on all the Storage Cluster nodes, run the following command on the deployment node:

```
# ceph-deploy install ceph-node{1,2,3,4}
```

You can now configure the Storage Cluster. See Section 1.8, “Configuring the Storage Cluster”.

### 1.8 Configuring the Storage Cluster

To configure the Storage Cluster, perform the following steps on the administration node:

1. Initialize Ceph monitoring and deploy a Ceph Monitor on one or more nodes in the Storage Cluster, for example:

```
# ceph-deploy mon create-initial
# ceph-deploy mon create ceph-node{2,3,4}
```

**Note**
For high availability, Oracle recommends that you configure at least three nodes as Ceph Monitors.

2. Gather the monitor keys and the OSD and MDS bootstrap keyrings from one of the Ceph Monitors, for example:

```
# ceph-deploy gatherkeys ceph-node3
```

3. Use the following command to prepare the back-end storage devices for each node in the Storage Cluster:

```
# ceph-deploy osd --zap-disk --fs-type fstype create node:device
```

**Note**
This command deletes all data on the specified device.

The supported file system types (`fstype`) are `btrfs` and `xfs`. 
Configuring the Storage Cluster

For example, prepare a btrfs file system as the back-end storage device on `/dev/sdb` for all nodes in a Storage Cluster:

```
# ceph-deploy osd --zap-disk --fs-type btrfs create ceph-node{1,2,3,4:sdb
```

4. When you have configured the Storage Cluster and established that it works correctly, re-enable SELinux in enforcing mode on each of the nodes where you previously disabled it and then reboot each node.

```
# sed -i '/SELINUX/s/disabled/enforcing/' /etc/selinux/config
# reboot
```

5. Restart, re-enable, and reconfigure the firewall service on each of the nodes where you previously disabled it.

For Oracle Linux 6:

a. Restart and re-enable the firewall service.

```
# service iptables start
# service ip6tables start
# chkconfig iptables on
# chkconfig ip6tables on
```

b. Allow access to TCP ports 6800 through 7300 that are used by the Ceph OSD, for example:

```
# iptables -A INPUT -i interface -p tcp -s network-address/netmask
    --match multiport --dports 6800:7300 -j ACCEPT
```

c. If a node runs Ceph Monitor, allow access to TCP port 6789, for example:

```
# iptables -A INPUT -i interface -p tcp -s network-address/netmask
    --dport 6789 -j ACCEPT
```

d. If a node is configured as an Object Gateway, allow access to port 7480 (or an alternate port that you have configured), for example:

```
# iptables -A INPUT -i interface -p tcp -s network-address/netmask
    --dport 7480 -j ACCEPT
```

e. Save the configuration:

```
# service iptables save
```

For Oracle Linux 7:

a. Restart and re-enable the firewall service.

```
# systemctl start firewalld
# systemctl enable firewalld
```

b. Allow access to TCP ports 6800 through 7300 that are used by the Ceph OSD, for example:

```
# firewall-cmd --permanent --zone=zone --add-port=6800-7300/tcp
```

c. If a node runs Ceph Monitor, allow access to TCP port 6789, for example:

```
# firewall-cmd --permanent --zone=zone --add-port=6789/tcp
```
1.9 Configuring an Object Gateway for OpenStack and Swift Access

If you want to enable access by OpenStack and Swift to Ceph, configure at least one Storage Cluster node as an Object Gateway.

For information on how to configure a node in a Storage Cluster as an Object Gateway, see [http://ceph.com/docs/master/install/install-ceph-gateway/](http://ceph.com/docs/master/install/install-ceph-gateway/).

1.10 Installing a Ceph Client

To install a Ceph Client:

1. Perform the following steps on the system that will act as a Ceph Client:
   a. If SELinux is enabled, disable it and then reboot the system.
      ```
      # sed -i '/SELINUX/s/enforcing/disabled/' /etc/selinux/config
      # reboot
      ```
   b. Stop and disable the firewall service.
      For Oracle Linux 6 or Oracle Linux 7 (where `iptables` is used instead of `firewalld`), enter:
      ```
      # service iptables stop
      # service ip6tables stop
      # chkconfig iptables off
      # chkconfig ip6tables off
      ```
      For Oracle Linux 7, enter:
      ```
      # systemctl stop firewalld
      # systemctl disable firewalld
      ```
   2. On the administration node of the Storage Cluster, copy the SSH key to the Ceph Client system, for example:
      ```
      # ssh-copy-id root@ceph-client
      ```
   3. On the deployment node (which is usually the same as the administration node), use `ceph-deploy` to install Ceph on the Ceph Client system, for example:
      ```
      # ceph-deploy install ceph-client
      ```
   4. On the administration node, copy the Ceph configuration file and the Ceph keyring to the Ceph Client system, for example:
      ```
      # ceph-deploy admin ceph-client
      ```

Note

This example assumes that you have configured entries for the Ceph Client system in DNS and/or in `/etc/hosts`. 
You can now configure a Block Device on the Ceph Client. See Section 1.12, “Configuring a Block Device on a Ceph Client”.

5. When you have established that the Ceph Client works with a Storage Cluster, re-enable SELinux in enforcing mode if you previously disabled it and then reboot the system.

```
# sed -i '/SELINUX/s/disabled/enforcing/' /etc/selinux/config
# reboot
```

### 1.11 Creating a Storage Pool

To create a storage pool for Block Devices in the OSD, use the following command:

```
# ceph osd pool create datastore 150 150
```

This example creates a pool named `datastore` with a placement group value of 150.

### 1.12 Configuring a Block Device on a Ceph Client

**Note**

Ensure that the Storage Cluster is active and healthy before configuring a block device.

To configure a Block Device on a Ceph Client:

1. Use the `rbd` command to create a Block Device image in the pool, for example:

```
# rbd create --size 4096 --pool datastore vol01
```

This example creates a 4096 MB volume named `vol01` in the `datastore` pool.

**Note**

If you do not specify a storage pool, `rbd` uses the default `rbd` pool:

```
# rbd create --size 4096 vol01
```

2. Use the `rbd` command to map the image to a Block Device, for example:

```
# rbd map vol01 --pool datastore
```

Ceph creates the Block Device under `/dev/rbd/pool/volume`.

The `rdb ls` command lists the images that you have mapped for a storage pool, for example:

```
# rdb ls -p datastore
vol01
```

You can make a file system on the Block Device and mount this file system on a suitable mount point, for example:

```
# mkfs.ext4 -m0 /dev/rbd/datastore/vol01
# mkdir /var/vol01
# mount /dev/rbd/datastore/vol01 /var/vol01
```

### 1.13 Known Issues

The following sections describe known issues in this release.
1.13.1 ceph-deploy Reports Errors

If `ceph-deploy` reports errors when run as a user with passwordless `sudo` privileges, run `visudo`, locate the `Defaults requiretty` entry, and either change this entry to `Defaults:ceph !requiretty` or comment out the entry.

1.13.2 ceph-deploy Reports an Error on Exit

Under Oracle Linux 6, `ceph-deploy` reports the error `Error in sys.exitfunc:` when it exits. This error is produced by Python's threading module and can safely be ignored. To prevent the error from being reported, set `CEPH_DEPLOY_TEST=YES` in the shell environment. (Bug ID 20755074)

1.13.3 ceph-deploy Python Errors on Oracle Linux 6

Under Oracle Linux 6, `ceph-deploy` reports errors such as the following in `syslog` and saves a Python crash dump:

```
timestamp server abrt: detected unhandled Python exception in '/usr/bin/ceph-deploy'
timestamp server abrt: New client connected
timestamp server abrt-server[PID]: Saved Python crash dump of pid PID to /var/spool/abrt/pyhook-timestamp-PID
timestamp server abrt: Directory 'pyhook-timestamp-PID' creation detected
timestamp server abrt: Duplicate: UUID
timestamp server abrt: DUP_OF_DIR: /var/spool/abrt/pyhook-timestamp-PID'
timestamp server abrt: Deleting problem directory pyhook-timestamp-PID (dup of pyhook-timestamp-PID')
timestamp server abrt: Sending an email...
timestamp server abrt: Email was sent to: root@localhost
```

Despite the apparent errors, the command succeeds. To prevent the errors from being reported and to stop the crash dump file from being written, set `CEPH_DEPLOY_TEST=YES` in the shell environment. (Bug ID 21089938)

1.13.4 ceph-radgosgw Service Fails to Start on Oracle Linux 7

If the owner of `/var/run/ceph` is `root`, the `ceph-radosgw` service fails to start on Oracle Linux 7, for example:

```
# systemctl start ceph-radosgw
Starting ceph-radosgw (via systemctl): Job for ceph-radosgw.service failed.
See 'systemctl status ceph-radosgw.service' and 'journalctl -xn' for details.
[FAILED]
```

The workaround is to run `visudo`, locate the `Defaults requiretty` entry, and either change this entry to `Defaults:ceph !requiretty` or comment out the entry. (Bug ID 21082202)

1.13.5 ceph-radgosgw Service Produces Error and Warning Messages when Starting on Oracle Linux 6

Under Oracle Linux 6, you might see the following warnings when you start the `ceph-radosgw` service:

```
# service ceph-radosgw start
Starting radosgw instance(s)...
/usr/bin/dirname: extra operand '-n'
```
Mounting a CephFS File System Produces a Spurious Error Message

1.13.6 Mounting a CephFS File System Produces a Spurious Error Message

Attempting to mount a CephFS file system on a client might produce the error `mount: error writing /etc/mtab: Invalid argument`. This error can safely be ignored.

Note

Although CephFS is provided, it is not supported and is not a feature of this technical preview.

(Bug ID 20469655)

1.13.7 sosreport Takes a Long Time to Run on a Storage Cluster Node

The `sosreport` command can take up to 45 minutes to run on a storage cluster node. (Bug ID 20523517)
Ceph Terminology

Block Device
A Ceph component that provides access to Ceph storage as a thinly provisioned block device. When an application writes to a Block Device, Ceph implements data redundancy and enhances I/O performance by replicating and striping data across the Storage Cluster.

Also known as a RADOS Block Device or RBD.

Ceph OSD
A Ceph component that provides access to an OSD.

Also known as a Ceph OSD Daemon.

Client
A host that can access the data stored in a Storage Cluster. A Ceph Client need not be a member node of a Storage Cluster.

Monitor (MON)
A Ceph component used for tracking active and failed nodes in a Storage Cluster.

Node
A system that is a member of a Storage Cluster.

Object Gateway
A Ceph component that provides a RESTful gateway that can use the Amazon S3 and OpenStack Swift compatible APIs to present OSD data to Ceph Clients, OpenStack, and Swift clients. An Object Gateway is configured on a node of a Storage Cluster.

Also known as a RADOS Gateway or RGW.

Object Storage Device (OSD)
Storage on a physical device or logical unit (LUN). Typically, data on an OSD is configured as a btrfs file system to take advantage of its snapshot features. However, other file systems such as XFS can also be used.

Storage Cluster
A Ceph component that stores MON and OSD data across cluster nodes.

Also known as a Ceph Object Store, RADOS Cluster, or Reliable Autonomic Distributed Object Store.