Load Balancing
Ceph Object Gateways

Deployment Guide
v1.0.0
1. About this Guide

This guide details the steps required to configure a load balanced Ceph Object Gateway environment utilizing Loadbalancer.org appliances. It covers the configuration of the load balancers and also any Ceph configuration changes that are required to enable load balancing.

For more information about initial appliance deployment, network configuration and using the Web User Interface (WebUI), please also refer to the relevant Administration Manual:

- v7 Administration Manual
- v8 Administration Manual

2. Loadbalancer.org Appliances Supported

All our products can be used for load balancing Ceph Object Gateway. The complete list of models is shown below:

<table>
<thead>
<tr>
<th>Discontinued Models</th>
<th>Current Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise R16</td>
<td>Enterprise R20</td>
</tr>
<tr>
<td>Enterprise VA R16</td>
<td>Enterprise MAX</td>
</tr>
<tr>
<td>Enterprise VA</td>
<td>Enterprise 10G</td>
</tr>
<tr>
<td>Enterprise R320</td>
<td>Enterprise 40G</td>
</tr>
<tr>
<td></td>
<td>Enterprise Ultra</td>
</tr>
<tr>
<td></td>
<td>Enterprise VA R20</td>
</tr>
<tr>
<td></td>
<td>Enterprise VA MAX</td>
</tr>
<tr>
<td></td>
<td>Enterprise AWS **</td>
</tr>
<tr>
<td></td>
<td>Enterprise AZURE **</td>
</tr>
<tr>
<td></td>
<td>Enterprise GCP **</td>
</tr>
</tbody>
</table>

* For full specifications of these models please refer to: [http://www.loadbalancer.org/products/hardware](http://www.loadbalancer.org/products/hardware)

** Some features may not be supported, please check with Loadbalancer.org support

3. Loadbalancer.org Software Versions Supported

- V76.4 and later

4. Ceph Software Versions Supported

- Ceph – all versions
5. Ceph

Ceph is a free and open-source object storage solution. It provides for the building of massively scalable and decentralised storage clusters, up to petabytes or exabytes of data, using standard commodity hardware.

At the heart of Ceph is its Reliable Autonomic Distributed Object Store (RADOS) which provides the underlying storage functionality. Ceph provides three interfaces for accessing its RADOS storage cluster: the Ceph File System (CephFS), which presents storage as a traditional POSIX-compatible file system; the Ceph Block Device (RBD), which allows storage to be mounted as a block device; and the Ceph Object Gateway (RGW), which provides S3 and Swift-compatible API access and is the subject of this document.

The Ceph project is backed by a variety of organisations, with contributions and broader community involvement notably from Canonical, CERN, Fujitsu, Intel, Red Hat, SanDisk, and SUSE. Commercial support and professional services for Ceph are available from Red Hat, as Red Hat Ceph Storage, SUSE, as SUSE Enterprise Storage, and Canonical, as “Ceph storage on Ubuntu”.

6. Load Balancing Ceph

Note: It’s highly recommended that you have a working Ceph environment with working Ceph Object Gateways first before implementing the load balancer.

Load Balancing & HA Requirements

For high availability and scalability, multiple Ceph Object Gateways must be deployed as part of a Ceph deployment. Additional gateways can be deployed over time to increase capacity to meet demand as needed.

Persistence (aka Server Affinity)

Ceph Object Gateways do not require session affinity at the load balancing layer.

Virtual Service (VIP) Requirements

To provide load balancing and HA for Ceph Object Gateways, one VIP is required:

- Ceph RGW

Optionally, an additional VIP can be used to make the Ceph Dashboard highly available from a single IP address, if the optional dashboard module is in use:

- Ceph Dashboard
Port Requirements

Ceph is extremely flexible and its user facing services can be configured to use HTTP and/or HTTPS over whichever ports are desired.

The following table is illustrative only: it shows a range of services using a mixture of default ports and likely real-world port values. A specific deployment may well use different ports. The services that are likely to be load balanced are:

<table>
<thead>
<tr>
<th>Port</th>
<th>Protocols</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>TCP/HTTP</td>
<td>Ceph Object Gateway Access</td>
</tr>
<tr>
<td>443</td>
<td>TCP/HTTPS</td>
<td>Secure Ceph Object Gateway Access (HTTPS)</td>
</tr>
<tr>
<td>8080</td>
<td>TCP/HTTP</td>
<td>Ceph Dashboard Access</td>
</tr>
<tr>
<td>8443</td>
<td>TCP/HTTPS</td>
<td>Secure Ceph Dashboard Access (HTTPS)</td>
</tr>
</tbody>
</table>

TLS/SSL Termination

Default/Recommended Option: No TLS/SSL Termination (TLS/SSL Pass-through)

If secure HTTPS services are in use, it is highly recommended to leave all TLS/SSL termination being performed on the Ceph Object Gateways. The gateway servers are best positioned to perform this function and, crucially, it's easy to provision additional gateway servers as needed to handle a growing computational load from increased HTTPS use.

Legacy Support Option: Performing TLS/SSL Termination On The Load Balancer

One exception to the above recommendation is when load balancing older Ceph deployments. Ceph version 11.0.1, Kraken, introduced support for TLS/SSL using the CivetWeb HTTP frontend on Ceph Object Gateways. Versions prior to that supported HTTP, plain text access to Ceph Object Gateways only.

To use HTTPS on an older deployment that doesn't have native HTTPS support, a TLS/SSL termination can be configured on the load balancer. This allows clients to connect to the load balancer using HTTPS and the load balancer can then send plain text HTTP to the Ceph Object Gateways, as normal.

To configure the load balancer in this way, first run through the instructions in section Appliance Configuration for Ceph Object Gateways – Using Layer 7 SNAT Mode (HAProxy) on page 21. Once complete, then proceed to run through the instructions in section 1 of the appendix, Legacy Support Option: Performing TLS/SSL Termination on the Load Balancer, on page 30.

Note: Layer 4 DR Mode is not compatible with the "legacy support", TLS/SSL termination option.

GSLB / Location Affinity

For multi-site Ceph deployments, it is possible to use the load balancer’s GSLB functionality to provide high availability and location affinity across multiple sites. Using this optional, DNS based feature, in the event that a site's Ceph Object Gateway service and/or load balancers are offline then local clients are automatically directed to a functioning Ceph Object Gateway service at another site.

A full explanation and instructions on setting up this optional feature can be found in section 2 of the appendix, Configuring GSLB / Location Affinity, on page 32.
7. Deployment Concept

Note: The load balancer can be deployed as a single unit, although Loadbalancer.org recommends a clustered pair for resilience & high availability. Please refer to section 3 in the appendix on page 36 for more details on configuring a clustered pair.

8. Load Balancer Deployment Methods

The load balancer can be deployed in 4 fundamental ways: Layer 4 DR mode, Layer 4 NAT mode, Layer 4 SNAT mode, and Layer 7 SNAT mode.

When load balancing Ceph Object Gateways, using layer 4 DR mode or layer 7 SNAT mode is recommended. It is also possible to use layer 4 TUN mode or layer 4 NAT mode in specific circumstances, however these load balancing methods come with many caveats and should only be considered if both layer 4 DR mode and layer 7 SNAT mode have been discounted.

Layer 4 DR mode and layer 7 SNAT mode are both described below and are used for the configurations presented in this guide. For configuring using DR mode please refer to the section starting on page 19, and for configuring using layer 7 SNAT mode refer to the section starting on page 21.

**Layer 4 DR Mode**

One-arm direct routing (DR) mode is a very high performance solution that requires little change to your existing infrastructure.
Note: Kemp, Brocade, Barracuda & A10 Networks call this Direct Server Return and F5 call it N-Path.

- DR mode works by changing the destination MAC address of the incoming packet to match the selected Real Server on the fly which is very fast.
- When the packet reaches the Real Server it expects the Real Server to own the Virtual Services IP address (VIP). This means that you need to ensure that the Real Server (and the load balanced application) respond to both the Real Servers own IP address and the VIP.
- The Real Server should not respond to ARP requests for the VIP. Only the load balancer should do this. Configuring the Real Servers in this way is referred to as Solving the ARP Problem. Please refer to section Configuring Ceph Object Gateways for Layer 4 DR Mode (LVS/DR) on page 11 for more information.
- On average, DR mode is 8 times quicker than NAT for HTTP, 50 times quicker for Terminal Services and much, much faster for streaming media or FTP.
- The load balancer must have an Interface in the same subnet as the Real Servers to ensure layer 2 connectivity required for DR mode to work.
- The VIP can be brought up on the same subnet as the Real Servers, or on a different subnet provided that the load balancer has an interface in that subnet.
- Port translation is not possible in DR mode i.e. having a different RIP port than the VIP port.
- DR mode is transparent, i.e. the Real Server will see the source IP address of the client.

**Layer 7 SNAT Mode**

Layer 7 SNAT mode uses a proxy (HAProxy) at the application layer. Inbound requests are terminated on the load balancer, and HAProxy generates a new request to the chosen Real Server. As a result, Layer 7 is a slower technique than DR or NAT mode at Layer 4. Layer 7 is typically chosen when either enhanced options such as SSL termination, cookie based persistence, URL rewriting, header insertion/deletion etc. are required, or when the network topology prohibits the use of the layer 4 methods.
This mode can be deployed in a one-arm or two-arm configuration and does not require any changes to the Real Servers. However, since the load balancer is acting as a full proxy it doesn't have the same raw throughput as the layer 4 methods.

The load balancer proxies the application traffic to the servers so that the source of all traffic becomes the load balancer.

- SNAT mode is a full proxy and therefore load balanced Real Servers do not need to be changed in any way
- Because SNAT mode is a full proxy any server in the cluster can be on any accessible subnet including across the Internet or WAN
- SNAT mode is not transparent by default, i.e. the Real Servers will not see the source IP address of the client, they will see the load balancers own IP address by default, or any other local appliance IP address if preferred (e.g. the VIP address), this can be configured per layer 7 VIP. If required, the clients IP address can be passed through either by enabling TProxy on the load balancer, or for HTTP, using X-forwarded-For headers. Please refer to chapter 6 in the administration manual for more details
- SNAT mode can be deployed using either a 1-arm or 2-arm configuration
Our Recommendation

The load balancing method to use depends on a number of factors. To help choose, presented below is a brief summary of each method.

Layer 4 DR Mode (LVS/DR)

This mode delivers the best raw performance and throughput. It requires making a few simple changes on the Ceph Object Gateways (see page 11 for details).

This is particularly well suited to read-intensive storage scenarios because the bandwidth of the return traffic is maximised by completely bypassing the load balancer. This is also known as ‘direct server return’.

This mode has the caveat that the load balancer must reside on the same network segment as the Ceph Object Gateways. If that condition cannot be met then layer 7 SNAT mode should be used.

Layer 7 SNAT Mode (HAProxy)

This is the simplest and most flexible mode to configure as no changes are required on the Ceph Object Gateways. The servers can be located anywhere in relation to the load balancer, provided that traffic can be routed between the load balancer and the servers.

In this mode, the load balancer uses HAProxy and acts as a full reverse proxy between the users and the Ceph Object Gateways.
9. Configuring Ceph for Load Balancing

Configuring Ceph Object Gateways for Layer 4 DR Mode (LVS/DR)

If Layer 4 DR mode is used then the 'ARP problem' must be solved. This involves configuring each Real Server (each Ceph Object Gateway in this context) to be able to receive traffic destined for the VIP address, and ensuring that each Real Server does not respond to ARP requests for the VIP address – only the load balancer should do this.

The steps below are for Linux servers. For instructions for FreeBSD servers, please refer to chapter 6 in our administration manual.

Step 1: Re-configure ARP On The Real Servers

To do this, add the following lines to /etc/sysctl.conf:

```bash
net.ipv4.conf.all.arp_ignore=1
net.ipv4.conf.ens192.arp_ignore=1
net.ipv4.conf.all.arp_announce=2
net.ipv4.conf.ens192.arp_announce=2
```

For example:

```
# sysctl settings are defined through files in
# /usr/lib/sysctl.d/, /run/sysctl.d/, and /etc/sysctl.d/.
# Vendors settings live in /usr/lib/sysctl.d/.
# To override a whole file, create a new file with the same in
# /etc/sysctl.d/ and put new settings there. To override
# only specific settings, add a file with a lexically later
# name in /etc/sysctl.d/ and put new settings there.
# For more information, see sysctl.conf(5) and sysctl.d(5).
net.ipv4.conf.all.arp_ignore=1
net.ipv4.conf.ens192.arp_ignore=1
net.ipv4.conf.all.arp_announce=2
net.ipv4.conf.ens192.arp_announce=2
```

Note: Adjust the commands shown above to suit the network configuration of your server, i.e. change ens192 to match the name of your server’s network interface.

Step 2: Re-configure DAD On The Real Servers (this Step Is Only Necessary If Using I Pv6 Addresses)

To do this add the following lines to /etc/sysctl.conf:
Step 3: Apply These Settings
Either reboot the Real Server or run the following command to apply these settings:

```
/sbin/sysctl -p
```

Step 4: Add The Virtual Service's IP Address To The Loopback Adapter
To temporarily add the VIP address, e.g. for an initial test, run the following command for the VIP address:

```
ip addr add dev lo <IPv4-VIP>/32
```

For IPv6 addresses use:

```
ip addr add dev lo <IPv6-VIP>/128
```
To make this new address permanent and persist across server reboots, add the command to an appropriate startup script, such as `etc/rc.local`. For example, on CentOS 7 / RHEL 7:

```bash
#!/bin/bash
# THIS FILE IS ADDED FOR COMPATIBILITY PURPOSES
#
# It is highly advisable to create own systemd services or udev rules
# to run scripts during boot instead of using this file.
#
# In contrast to previous versions due to parallel execution during boot
# this script will NOT be run after all other services.
#
# Please note that you must run 'chmod +x /etc/rc.d/rc.local' to ensure
# that this script will be executed during boot.

touch /var/lock/subsys/local
ip addr add dev lo 172.26.11.231/32

"/etc/rc.local" 15L, 510C 15,35 All
```

As described in the help text, on CentOS 7 / RHEL 7 it is necessary to also execute

```
chmod +x /etc/rc.d/rc.local
```

to make the rc.local script executable during boot, otherwise the VIP address will not be added at boot.
Note: As an alternative to using a startup script such as rc.local (which, on some Linux distributions, may be considered a legacy option or is offered “for compatibility purposes” only), the VIP address can be added to the loopback adaptor by modifying the OS appropriate NIC configuration file.

As an example, on CentOS 7 / RHEL 7 the loopback adaptor configuration file at /etc/sysconfig/network-scripts/ifcfg-lo can be appended to in order to add the additional IP address, e.g. by appending:

```
IPADDR1=172.26.11.231
NETMASK1=255.255.255.255
```

The network service would then need to be restarted to put the new configuration and IP address into use, e.g. by executing `systemctl restart network`

### Configuring the Dashboard Module for Load Balancing

The Ceph Dashboard module is an optional component of Ceph. If desired, an additional virtual service can be used to make the Ceph Dashboard highly available from a single IP address.

Two changes should be made to the dashboard module before attempting to load balance the service.

#### Disable The Redirection

The default behaviour of passive manager nodes is to send an HTTP 303 response status code when a client attempts to access the dashboard from them. When the dashboard service is being load balanced, there is a possibility that a connecting client could be sent an unresolvable URL redirect in the event that a fail over from one manager to another is in progress. As such, the default redirection behaviour should be disabled when the dashboard service is being load balanced.

From any Ceph node in the cluster, execute the following command (sudo privileges may be required):

```
ceph config set mgr mgr/dashboard/standby_behaviour "error"
```
Configure The Error Status Code

Once redirection has been disabled, passive manager nodes will send an HTTP 500 'internal server error' status code if a client attempts to access the dashboard from them. It is recommended by the Ceph project, and is good practice, to use the more descriptive HTTP 503 'service unavailable' status code.

From any Ceph node in the cluster, execute the following command (sudo privileges may be required):

```bash
ceph config set mgr mgr/dashboard/standby_error_status_code 503
```
10. Loadbalancer.org Appliance – the Basics

Virtual Appliance Download & Deployment
A fully featured, fully supported 30 day trial is available if you are conducting a PoC (Proof of Concept) deployment. The VA is currently available for VMware, Virtual Box, Hyper-V, KVM and XEN and has been optimized for each Hypervisor. By default, the VA is allocated 1 CPU, 2GB of RAM and has an 8GB virtual disk. The Virtual Appliance can be downloaded [here](#).

Note: The same download is used for the licensed product, the only difference is that a license key file (supplied by our sales team when the product is purchased) must be applied using the appliance’s WebUI.

Note: Please refer to the [Administration Manual](#) and the ReadMe.txt text file included in the VA download for more detailed information on deploying the VA using various Hypervisors.

Initial Network Configuration
The IP address, subnet mask, default gateway and DNS settings can be configured in several ways as detailed below:

**Method 1 - Using the Network Setup Wizard at the console**
After boot up, follow the instructions on the console to configure the IP address, subnet mask, default gateway and DNS settings.

**Method 2 - Using the WebUI**
Using a browser, connect to the WebUI on the default IP address/port: [https://192.168.2.21:9443](https://192.168.2.21:9443)
To set the IP address & subnet mask, use: *Local Configuration > Network Interface Configuration*
To set the default gateway, use: *Local Configuration > Routing*
To configure DNS settings, use: *Local Configuration > Hostname & DNS*
Accessing the Web User Interface (WebUI)
The WebUI can be accessed via HTTPS at the following URL: https://192.168.2.21:9443/lbadmin
* Note the port number → 9443

(replace 192.168.2.21 with the IP address of your load balancer if it's been changed from the default)

Login using the following credentials:

**Username:** loadbalancer
**Password:** loadbalancer

Note: To change the password, use the WebUI menu option: Maintenance > Passwords.

Once logged in, the WebUI will be displayed as shown on the following page:
HA Clustered Pair Configuration

Loadbalancer.org recommend that load balancer appliances are deployed in pairs for high availability. In this guide a single unit is deployed first, adding a secondary slave unit is covered in section 3 of the appendix on page 36.
11. Appliance Configuration for Ceph Object Gateways – Using Layer 4 DR Mode (LVS/DR)

The following instructions assume that the steps from section Configuring Ceph Object Gateways for Layer 4 DR Mode (LVS/DR) on page 11 have already been followed.

Configuring VIP 1 – Ceph Object Gateways

Configuring The Virtual Service (VIP)

1. Using the web user interface, navigate to Cluster Configuration > Layer 4 – Virtual Services and click on Add a new Virtual Service
2. Define the Label for the virtual service as required, e.g. Ceph_RGW
3. Set the Virtual Service IP Address field to the required IP address, e.g. 172.26.11.231
4. Set the Ports field to the ports that are in use by the Ceph Object Gateways, e.g. 80,443
5. Leave the Protocol set to TCP
6. Leave the Forwarding Method set to Direct Routing
7. Click Update to create the virtual service

8. Click Modify next to the newly created VIP
9. Set the Balance Mode to Weighted Round Robin
10. Ensure that the Persistence Enable checkbox is not checked
11. Click Update
Defining The Real Servers (RIPs)

1. Using the web user interface, navigate to Cluster Configuration > Layer 4 – Real Servers and click on Add a new Real Server next to the newly created VIP
2. Define the Label for the real server as required, e.g. RGW_Node1
3. Set the Real Server IP Address field to the required IP address, e.g. 172.26.11.220
4. Click Update
5. Repeat these steps to add the remaining Ceph Object Gateways

<table>
<thead>
<tr>
<th>Layer 4 Add a new Real Server - Ceph_RGW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
</tr>
<tr>
<td>Real Server IP Address</td>
</tr>
<tr>
<td>Weight</td>
</tr>
<tr>
<td>Minimum Connections</td>
</tr>
<tr>
<td>Maximum Connections</td>
</tr>
</tbody>
</table>

Configuring VIP 2 – Ceph Dashboard

A layer 7 solution (which uses HAProxy) should be used for this optional virtual service. To configure it, follow the instructions in section Configuring VIP 2 – Ceph Dashboard on page 23 followed by the instructions that immediately follow it, Finalizing the Layer 7 Configuration, on page 25.
12. Appliance Configuration for Ceph Object Gateways – Using Layer 7 SNAT Mode (HAProxy)

**Enabling Multithreaded Load Balancing**

For the full layer 7 load balancing scenario (not necessary if using a layer 7 virtual service for the dashboard element only), the Loadbalancer.org appliance should be configured to actively use multiple CPU cores for the load balancing process. This is required to achieve the high level of performance and throughput required when load balancing a deployment of Ceph Object Gateways at layer 7.

Note: A virtual host should be allocated a minimum of 4 vCPUs.

To enable multithreaded mode from the WebUI:

1. Navigate to Cluster Configuration > Layer 7 - Advanced Configuration
2. Check the Enable Multithreading checkbox
3. Set Number of Threads as appropriate, e.g. 3

Note: We recommend using a number of threads equal to ‘the number of CPU cores - 1’. For example: on a 4 core CPU system, 3 threads are recommended; on an 8 core CPU system, 7 threads are recommended.

4. Click Update to apply the changes

**Configuring VIP 1 – Ceph Object Gateways**

**Configuring The Virtual Service (VIP)**

1. Using the web user interface, navigate to Cluster Configuration > Layer 7 – Virtual Services and click on Add a new Virtual Service
2. Define the Label for the virtual service as required, e.g. Ceph_RGW
3. Set the Virtual Service IP Address field to the required IP address, e.g. 172.26.11.232
4. Set the **Ports** field to the ports that are in use by the Ceph Object Gateways, e.g. **80,443**
5. Set the **Layer 7 Protocol** to **TCP Mode**
6. Click **Update** to create the virtual service

![Layer 7 - Add a new Virtual Service](image)

7. Click **Modify** next to the newly created VIP
8. Set the **Balance Mode** to **Weighted Round Robin**
9. Set **Persistence Mode** to **None**
10. Click **Update**

**Defining The Real Servers (RIPs)**

1. Using the web user interface, navigate to **Cluster Configuration > Layer 7 – Real Servers** and click on **Add a new Real Server** next to the newly created VIP
2. Define the **Label** for the real server as required, e.g. **RGW_Node1**
3. Set the **Real Server IP Address** field to the required IP address, e.g. **172.26.11.220**
4. Click **Update**
5. Repeat these steps to add the remaining Ceph Object Gateways
Configuring VIP 2 – Ceph Dashboard

The Ceph Dashboard module is an optional component of Ceph. In a deployment with multiple manager nodes, only one manager node is active at a given time. Only the active manager node hosts the web-based dashboard service.

If desired, an additional virtual service can be used on the load balancer to ensure that the Ceph Dashboard is highly available and always accessible from the same IP address.

The dashboard should have been configured for load balancing before following the instructions presented below (see section Configuring the Dashboard Module for Load Balancing on page 14).

Configuring The Virtual Service (VIP)

1. Using the web user interface, navigate to Cluster Configuration > Layer 7 – Virtual Services and click on Add a
new Virtual Service

2. Define the Label for the virtual service as required, e.g. Ceph_Dashboard

3. Set the Virtual Service IP Address field to the required IP address, e.g. 172.26.11.233

4. Set the Ports field to the ports that are in use by the Ceph Dashboard, e.g. 8080,8443

5. Set the Layer 7 Protocol to TCP Mode

6. Click Update to create the virtual service

7. Click Modify next to the newly created VIP

8. In the Health Checks section click Advanced to expand the menu

9. Set Health Checks as appropriate for the dashboard configuration. For example:
   - If the dashboard service is configured to use HTTP then a valid option is to select Negotiate HTTP (GET) and set the Check Port as required, e.g. 8080
   - If the dashboard service is configured to use HTTPS then a valid option is to select Negotiate HTTPS (GET) and set the Check Port as required, e.g. 8443

10. In the Other section click Advanced to expand the menu

11. Check the Timeout checkbox

12. Set Client Timeout to 50s (for 50 seconds)

13. Set Real Server Timeout to 450s

14. Click Update

Defining The Real Servers (RIPs)

1. Using the web user interface, navigate to Cluster Configuration > Layer 7 – Real Servers and click on Add a new Real Server next to the newly created VIP
2. Define the Label for the real server as required, e.g. Manager_Node1
3. Set the Real Server IP Address field to the required IP address, e.g. 172.26.11.200
4. Click Update
5. Repeat these steps to add the remaining manager nodes

![Layer 7 Add a new Real Server - Ceph_Dashboard](image)

Note: Unlike with most load balancing deployments, it is required to add all manager nodes to the dashboard virtual service. This is because any manager node could be the single active node at a given time. For example, if a manager node becomes the active node but is not defined under the dashboard virtual service then the virtual service will fail.

**Finalizing the Layer 7 Configuration**

To apply the new settings, HAProxy must be reloaded as follows:

1. Using the WebUI, navigate to: Maintenance > Restart Services and click Reload HAProxy
13. Testing & Verification

Using System Overview

The System Overview can be viewed in the WebUI. It shows a graphical view of all VIPs & RIPv (i.e. the Ceph nodes) and shows the state/health of each server as well as the state of the cluster as a whole.

The example below shows a layer 4 DR mode configuration load balancing a pair of Ceph Object Gateways, where both nodes are healthy and available to accept connections, and a layer 7 virtual service for the Ceph Dashboard, with the active manager node showing as available to accept connections:

![System Overview](image)

The example below shows a layer 7 SNAT mode configuration load balancing a pair of Ceph Object Gateways, where both nodes are healthy and available to accept connections, and a layer 7 virtual service for the Ceph Dashboard, with the active manager node showing as available to accept connections:
Testing Dashboard Fail Over

It is possible to issue a command to intentionally cause the active manager node to fail. This will cause one of the other manager nodes to become active and start serving the Ceph Dashboard service. From the load balancer’s System Overview page, the active manager node should then turn ‘red’/offline and another manager node should become ‘green’/healthy.

From any Ceph node in the cluster, execute the following command to trigger a manager node failure (sudo privileges may be required):

```
ceph mgr fail <hostnameOfActiveManager>
```

In the following example, the manager daemons all reside on monitor nodes. A failure is triggered on the active manager daemon, which resides on monitor node 1:

```
[cephDeployUser@mon-node1 ~]$ sudo ceph mgr fail mon-node1
[cephDeployUser@mon-node1 ~]$
```
It is then observed that another manager node has become active:

As a final test, the dashboard is successfully logged into using the VIP address, which is 172.26.11.233 in this example:

14. Technical Support
For more details about configuring the appliance and assistance with designing your deployment please don't hesitate to contact the support team using the following email address: support@loadbalancer.org.

15. Further Documentation
16. Conclusion
Loadbalancer.org appliances provide a very cost effective solution for highly available load balanced Ceph environments with Ceph Object Gateways.
17. Appendix

1 – Legacy Support Option: Performing TLS/SSL Termination on the Load Balancer

Ceph versions prior to 11.0.1, Kraken, supported HTTP access to Ceph Object Gateways only. To use HTTPS on an older deployment which doesn't have native HTTPS support, a TLS/SSL termination can be configured on the load balancer. This allows clients to connect to the load balancer using HTTPS, and the load balancer can then send plain text HTTP to the Ceph Object Gateways.

The following instructions assume that the steps from section Appliance Configuration for Ceph Object Gateways – Using Layer 7 SNAT Mode (HAProxy) on page 21 have already been followed.

Uploading The Certificate

A self signed public certificate is included on the load balancer. This can be used for handling TLS/SSL connections, however, connecting clients will need to manually intervene to accept the self signed certificate when connecting to the load balanced HTTPS service. This is often acceptable for internal deployments. For public facing deployments, it is strongly advised to obtain a public certificate signed by a well known and trusted certificate authority.

Once obtained, the appropriate public certificate, including both the private key and public certificate parts, must be uploaded to the load balancer for TLS/SSL termination to work.

Detailed information on creating PEM certificate files and converting between certificate formats is presented in our Administration Manual: http://pdfs.loadbalancer.org/loadbalanceradministrationv8.pdf

The process for uploading a certificate is as follows:

1. Using the web user interface, navigate to Cluster Configuration > SSL Certificate and click on Add a new SSL Certificate
2. Press the Upload prepared PEM/PFX file radio button
3. Define the Label for the certificate as required. It may make sense to use the domain that the certificate is associated to, e.g. ceph.mysite.org
4. Click on Browse and select the appropriate PEM or PFX style certificate
5. If uploading a PFX certificate, enter the certificate's password in the PFX File Password field
6. Click Upload certificate

Modifying The Existing Ceph Object Gateways Virtual Service

1. Using the web user interface, navigate to Cluster Configuration > Layer 7 – Virtual Services and click Modify next to the Ceph Object Gateways virtual service (Ceph_RGW in the example presented in this document)
2. Set the Ports field to the port that is in use by the Ceph Object Gateways for the HTTP service, e.g. 80
3. Click Update

Creating The TLS/SSL Termination

1. Using the web user interface, navigate to Cluster Configuration > SSL Termination and click on Add a new Virtual Service
2. From the Associated Virtual Service drop-down list, select the Ceph Object Gateways service that was created previously, e.g. Ceph_RGW
3. Set the Virtual Service Port field to the port that will be used for the HTTPS service, e.g. 443
4. If using an uploaded public certificate, i.e. not the default self signed certificate, from the SSL Certificate drop-down list, select the certificate for the service in question, which in this example is **ceph.mysite.org**

5. Click **Update** to create the TLS/SSL termination service

![SSL Termination - Add a new Virtual Service](image)

**Finalizing The Configuration**

To apply the new settings, HAProxy and stunnel must both be reloaded as follows:

1. Using the WebUI, navigate to: **Maintenance > Restart Services** and click **Reload STunnel**
2. Using the WebUI, navigate to: **Maintenance > Restart Services** and click **Reload HAProxy**
2 – Configuring GSLB / Location Affinity

Conceptual Overview

For multi-site Ceph deployments, it is possible to use the load balancer’s GSLB functionality to provide high availability and location affinity across multiple sites.

- Clients across multiple sites can use the same fully qualified domain name to access the Ceph Object Gateway service
- Under normal operation, clients are directed to their site’s local cluster of Ceph Object Gateways
- In the event that a site’s Ceph Object Gateway service and/or load balancers are offline then local clients are automatically directed to a functioning cluster of Ceph Object Gateways at another site

For the sake of simplicity, the diagram presented below shows a two site setup. The principle can be extended to encompass as many sites as desired.
**Explanation:**

- **Start:** A client tries to access the Ceph Object Gateway service using the S3 protocol. To do this, the client uses the service's fully qualified domain name, in this example `s3-region1.domain.tld`
- The client sends a DNS query for `s3-region1.domain.tld` to its local DNS server
- The local site's DNS server has the domain `s3-region1.domain.tld` delegated to the load balancers
- The DNS server sends a delegated DNS query for `s3-region1.domain.tld` to one of the load balancers
- The load balancer that received the delegated DNS query replies to the DNS server by serving up the appropriate, local VIP address. For example, if the query originated from the 10.0.0.0/24 subnet then the VIP in that subnet is served up. Likewise, if the query originated from the 172.16.0.0/24 subnet then the VIP in that subnet is served up. As such, clients are always directed to their local, on-site Ceph Object Gateway cluster, provided that the on-site instances are online and available
- The DNS server sends the delegated DNS answer to the client
- **Finish:** The client connects to the S3 service at `s3-region1.domain.tld` by using the local VIP address

---

**Note:** In the event that the cluster of Ceph Object Gateways and/or load balancers at one site should completely fail then local clients will be directed to the cluster of Ceph Object Gateways at the other site and the service will continue to be available.

This style of multi-site failover is possible because the load balancers' GSLB functionality continuously health checks the service at each site. When the service at a site is observed to be unavailable then that site's IP address is no longer served when responding to DNS queries.

---

**Handling Multiple Subdomains**

Object storage related DNS configurations often use various DNS subdomains, for example:

- `s3-<region/location>.domain.tld` (e.g. `s3-region1.domain.tld`)

Due to the complexity of DNS delegation, it is recommended to delegate a single subdomain using the GSLB service and then use CNAME records to point everything else at the delegated subdomain. This makes it much easier to configure many DNS entries, including more complex options such as using wildcard entries, for example `*.s3-region1.domain.tld`.

---

**Appliance Configuration**

The GSLB service must be configured on the **master** load balancer appliance at each site. The GSLB configuration must be identical across all sites. Configuration takes place in two locations in the WebUI:

- **Polaris config:** the main configuration for the Polaris service which handles GSLB
- **Topology config:** defines the network topology, mapping network subnets to sites

---

**Polaris Config**

Using the web user interface of the master appliance, navigate to **Cluster Configuration > GSLB Configuration** and select **Polaris Config** from the drop-down list.

If GSLB has never been configured then the default example configuration will be displayed in the text box.
What follows is an example configuration which covers three example sites. It can be followed as a template to replace the default example and can be used as a basis for creating a deployment-specific configuration.

The elements presented in **boldface** should be customised to reflect the Ceph deployment in question. Stanzas should be added or removed under the ‘members’ section as needed to reflect all sites that are to be used in the deployment.

Be sure to click the **Update** button when finished to write the configuration to disk.

```
Note: Indentation is important and **must** be preserved, otherwise the underlying Polaris service will throw an error. As many PDF readers discard leading whitespace, the below example should **not** be directly copied and pasted into the Polaris configuration text box. The preexisting, default configuration should be reworked as necessary, with its indentation rules followed and preserved.

```

globalnames:
  s3-region1.domain.tld:
    pool: rgw-nodes
    ttl: 5

pools:
  rgw-nodes:
    monitor: tcp
    monitor_params:
      port: 80
      lb_method: twrr
      fallback: any
    members:
      - ip: 10.0.0.2
        name: rgw-vip_dc1
        weight: 1
      - ip: 172.16.0.2
        name: rgw-vip_dc2
        weight: 1
      - ip: 192.168.1.2
        name: rgw-vip_dc3
        weight: 1

Topology Config

Using the web user interface of the master appliance, navigate to Cluster Configuration > GSLB Configuration and select **Topology Config** from the drop-down list.

If GSLB has never been configured then the default example topology configuration will be displayed in the text box.

What follows is an example topology configuration which covers the same three example sites from the example configuration in the previous section. It can be copied and pasted to replace the default topology example and can be used as a basis for creating a deployment-specific configuration.

The titles and IP subnets should be customised to reflect the Ceph deployment in question. Stanzas should be added or removed as needed to reflect all sites that are to be used in the deployment.

Be sure to click the **Update** button when finished to write the configuration to disk.
datacenter1:
- 10.0.0.0/24

datacenter2:
- 172.16.0.0/24

datacenter3:
- 192.168.0.0/18

Applying the Configuration

If updating the configuration of an existing GSLB configuration then press the Reload GSLB button when prompted. If setting up the GSLB service for the first time then the service must undergo a full restart to be enabled and for the configuration to be applied. To do this, using the web user interface, navigate to Maintenance > Restart Services and click the Restart GSLB button. If using an HA pair of load balancers at a site then this procedure must also be carried out on the slave appliance. Once this procedure been carried out it never needs to be done again; the GSLB service only needs to be reloaded in the future when prompted, for example following a configuration change, something that only needs to be carried out on the master appliance.

DNS Server Configuration

Once the GSLB service has been configured on the master load balancer at every site, ensuring that the configuration is identical across all sites, the DNS server at each site must then be configured for GSLB.

The DNS server at each site must be configured to delegate DNS requests for the subdomain in question to the load balancers; the load balancers' GSLB services will serve the appropriate IP addresses to the DNS servers. Using the example presented throughout this appendix, the DNS server at each site would be configured with a delegation for the domain s3-region1.domain.tld. The domain would be delegated to every load balancer across every site, which provides multi-site redundancy.

The exact steps for creating a DNS delegation vary between different DNS servers and are outside the scope of this document. For further information, a blog post that walks through creating a DNS delegation on a Microsoft DNS server in the context of setting up GSLB on our appliance can be found at https://www.loadbalancer.org/blog/loadbalancer-org-releases-a-gslb/ (see the section titled “Delagating your subdomain to your GSLB’s using Microsoft’s DNS Server”).
3 – Clustered Pair Configuration – Adding a Slave Unit

If you initially configured just the master unit and now need to add a slave - our recommended procedure, please refer to the relevant section below for more details:

Note: A number of settings are not replicated as part of the master/slave pairing process and therefore must be manually configured on the slave appliance. These are listed below:

- Hostname & DNS settings
- Network settings including IP addresses, bonding configuration and VLANs
- Routing configuration including default gateways and static routes
- Date & time settings
- Physical – Advanced Configuration settings including Internet Proxy IP address & port, Firewall table size, SMTP relay and Syslog server
- SNMP settings
- Graphing settings
- Firewall Script & Firewall Lockdown Script settings
- Software updates

Version 7:
Please refer to Chapter 8 – Appliance Clustering for HA in the v7 Administration Manual.

Version 8:
To add a slave node – i.e. create a highly available clustered pair:

- Deploy a second appliance that will be the slave and configure initial network settings
- Using the WebUI, navigate to: Cluster Configuration > High-Availability Configuration
• Specify the IP address and the loadbalancer users password (the default is 'loadbalancer') for the slave (peer) appliance as shown above

• Click Add new node

• The pairing process now commences as shown below:

![Create a Clustered Pair](image)

• Once complete, the following will be displayed:

![High Availability Configuration - Master](image)

• To finalize the configuration, restart heartbeat and any other services as prompted in the blue message box at the top of the screen

Note: Clicking the Restart Heartbeat button on the master appliance will also automatically restart heartbeat on the slave appliance.

Note: Please refer to chapter 9 – Appliance Clustering for HA in the Administration Manual for more detailed information on configuring HA with 2 appliances.
## 18. Document Revision History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Change</th>
<th>Reason for Change</th>
<th>Changed By</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0.0</td>
<td>23 March 2020</td>
<td>Initial version</td>
<td></td>
<td>AH</td>
</tr>
</tbody>
</table>
About Loadbalancer.org

Loadbalancer.org's mission is to ensure that its clients' businesses are never interrupted. The load balancer experts ask the right questions to get to the heart of what matters, bringing a depth of understanding to each deployment. Experience enables Loadbalancer.org engineers to design less complex, unbreakable solutions - and to provide exceptional personalized support.

United Kingdom
Loadbalancer.org Ltd.
Compass House, North Harbour Business Park, Portsmouth, PO6 4PS
UK:+44 (0) 330 380 1064
sales@loadbalancer.org
support@loadbalancer.org

Canada
Loadbalancer.org Appliances Ltd.
300-422 Richards Street, Vancouver, BCV6B 2Z4, Canada
TEL:+1 302.213.0122
sales@loadbalancer.org
support@loadbalancer.org

United States
Loadbalancer.org, Inc.
4550 Linden Hill Road, Suite 201 Wilmington, DE 19808, USA
TEL: +1 833.274.2566
sales@loadbalancer.org
support@loadbalancer.org

Germany
Loadbalancer.org GmbH
Tengstraße 2780798, München, Germany
TEL: +49 (0)89 2000 2179
sales@loadbalancer.org
support@loadbalancer.org

© Copyright Loadbalancer.org • www.loadbalancer.org