OpenStack Load Balancers
Developer Guide

OpenStack Load Balancers Developer Guide

API v1.1

Publication date 2011-04-19

Abstract

This document is intended for software developers interested in developing applications using the OpenStack Load Balancers Application Program Interface (API).

Table of Contents

List of Tables

List of Examples

Chapter 1. Overview

1.1. Intended Audience

This guide is intended for software developers who want to create applications using the OpenStack Load Balancers API. It assumes the reader has a general understanding of load balancing concepts and is familiar with:

- ReSTful web services
- HTTP/1.1 conventions
- JSON and/or XML serialization formats

1.2. Document Change History

This version of the Developer Guide replaces and obsoletes all previous versions. The most recent changes are described in the table below:

Revision Date	Summary of Changes
Apr. 19, 2011	• Added details to support initial GA release.
Mar. 2, 2011	• Revised code samples and formatting to address initial beta feedback.
Feb. 23, 2011	• Initial release for public beta.

Chapter 2. Concepts

To use OpenStack Load Balancers API effectively, you should understand several key concepts:

2.1. Load Balancer

A load balancer is a logical device. It is used to distribute workloads between multiple back-end systems or services called nodes, based on the criteria defined as part of its configuration.

2.2. Virtual IP

A virtual IP is an Internet Protocol (IP) address configured on the load balancer for use by clients connecting to a service that is load balanced. Incoming connections and requests are distributed to back-end nodes based on the configuration of the load balancer.

2.3. Node

A node is a back-end device providing a service on a specified IP and port.

2.4. Health Monitor

A health monitor is a feature of each load balancer. It is used to determine whether or not a back-end node is usable for processing a request. The load balancing service supports two types of health monitors: passive and active.

2.4.1. Passive Health Monitor

By default, all load balancing configurations utilize a passive health monitor, which is the default monitoring and does not require configuration from the user. If the passive health monitoring determines that a node is down, unreachable or malfunctioning, it puts the node in an OFFLINE state and stops sending traffic to it.

2.4.2. Active Health Monitor

Active health monitoring is a technique that is configured by users. It uses synthetic transactions executed at periodic intervals to determine the condition of a node. When active monitoring is configured, it takes over the monitoring of the nodes, and passive monitoring is disabled. Conversely, when active monitoring configuration is removed by the user, passive monitoring is re-enabled for the nodes of the load balancer

The active health monitor can use one of three types of probes:

- CONNECT
- HTTP
- HTTPS

These probes are executed at configured intervals; in the event of a failure, the node status changes to `OFFLINE` and the node will not receive traffic. If, after running a subsequent test, the probe detects that the node has recovered, then the node's status is changed to `ONLINE` and it is capable of servicing requests.

2.5. Session Persistence

Session persistence is a feature of the load balancing service. It attempts to force subsequent connections to a service to be redirected to the same node as long as it is online.

2.6. Connection Logging

The connection logging feature allows for retrieving access logs (for HTTP-based protocol traffic) or connection and transfer logs (for all other traffic)

Chapter 3. General API Information

Sections in this chapter describe operations and guidelines that are common to all OpenStack APIs, not specific to the Load Balancing API.

3.1. Authentication

Load Balancing API will use the standard defined by OpenStack for authentication. Once this is finalized, this section will describe how users can authenticate and obtain a token from the OpenStack Authentication service, and how they can present this token using the Load Balancing API.

Refer to current status of OpenStack Authentication at http://wiki.openstack.org/openstack-authn for detailed information

3.2. Service Access/Endpoints

Your service provider will publish the endpoints that you can use to connect to their LB service.

3.3. Request/Response Types

The Cloud Load Balancers API supports both the JSON and XML data serialization formats. The request format is specified using the `Content-Type` header and is required for operations that have a request body. The response format can be specified in requests using either the `Accept` header or adding an `.xml` or `.json` extension to the request URI. Note that it is possible for a response to be serialized using a format different from the request. If no response format is specified, JSON is the default. If conflicting formats are specified using both an `Accept` header and a query extension, the query extension takes precedence.

Table 3.1. JSON and XML Response Formats

Format	Accept Header	Query Extension	Default
JSON	application/json	.json	Yes
XML	application/xml	.xml	No

3.4. Content Compression

Request and response body data may be encoded with gzip compression to accelerate interactive performance of API calls and responses. This is controlled using the `Accept-Encoding` header in the request from the client and indicated by the `Content-Encoding` header in the server response. Unless the header is explicitly set, encoding defaults to disabled.

Table 3.2. Encoding Headers

Header Type	Name	Value
HTTP/1.1 Request	`Accept-Encoding`	`gzip`
HTTP/1.1 Response	`Content-Encoding`	`gzip`

3.5. Persistent Connections

By default, the API supports persistent connections via `HTTP/1.1` keepalives. All connections will be kept alive unless the connection header is set to close.

Note

The server may close the connection at any time and clients should not rely on this behavior.

3.6. Paginated Collections

To reduce load on the service, list operations will return a maximum of 100 items at a time. To navigate the collection, the `limit` and `marker` parameters (for example, `?limit=50&marker=1`) can be set in the URI. If a marker beyond the end of a list is given, an empty list is returned. Note that list operations never return 404 (itemNotFound) faults.

3.7. Limits

Accounts may be preconfigured set of thresholds (or limits) to manage capacity and prevent abuse of the system. The system recognizes two kinds of limits: rate limits and absolute limits. Rate limits are thresholds that are reset after a certain amount of time passes. Absolute limits are fixed.

3.7.1. Rate Limits

We specify rate limits in terms of both a human-readable wild-card URI and a machine-processable regular expression. The regular expression boundary matcher '^' takes effect after the root URI path. For example, the regular expression `^/v1.1/1234/loadbalancers` would match the bolded portion of the following URI: `https://loadbalancers.api.openstack.org` **`/v1.1/1234/loadbalancers`**.

Table 3.3. Default Rate Limits

Verb	URI	RegEx	Default Limit
GET	/v1.1/*	^/1.1/.*	5/second
POST	/v1.1/*	^/1.1/.*	2/second
POST	/v1.1/*	^/1.1/.*	25/minute
PUT	/v1.1/*	^/1.1/.*	5/second
DELETE	/v1.1/*	^/1.1/.*	2/second

Rate limits are applied in order relative to the verb, going from least to most specific. For example, although the threshold for **POST** to /v1.1/* is 25 per minute, one cannot **POST** to /v1.1/* more than 2 times per second because the rate limit for any **POST** is 2 per second. In the event you exceed the thresholds established for your account, a 413 (Rate Control) HTTP response will be returned with a `Retry-After` header to notify the client when they can attempt to try again.

3.7.2. Absolute Limits

Absolute limits are specified as name/value pairs. Then name of the absolute limit uniquely identifies the limit within a deployment. For example `maxNodesPerLoadbalancer` identifies the total number of nodes that may be associated with a given load balancer.

The following table shows some of these limits and example values:

Name	Value	Description
maxLoadBalancers	20	Maximum loadbalancers that can be created for this account

Name	Value	Description
maxNodesPerLoadBalancer	5	Maximum nodes allowed per load balancer
maxVIPsperLoadBalancer	2	Maximum VIPs allowed per load balancer
maxDaysForDeletedLoadBalancers	15	The maximum number of days that deleted loadbalancers can be queried by the user
maxLoadBalancerNameLength	15	Maximum length of a load balancer name

3.7.3. Determining Limits Programmatically

Applications can programmatically determine current account limits using the `/limits` URI as follows:

Verb	URI	Description
GET	/limits	Return the current limits for the account.

Normal Response Code(s): 200

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

Example 3.1. List Limits Response: XML

```
        <limits xmlns="http://docs.openstack.org/common/api/v1.1">
    <rates>
        <rate uri="/v1.1/*" regex="^/1.1/.*">
            <limit
                verb="GET"
                value="600000"
                remaining="426852"
                unit="HOUR"
                next-available="2011-02-22T19:32:43.835Z"/>
        </rate>
    </rates>
    <absolute>
        <limit name="maxNodesPerLoadBalancer" value="5"/>
        <limit name="maxVIPsperLoadBalancer" value="2"/>
  </absolute>
</limits>
```

Example 3.2. List Limits Response: JSON

```
        {
    "limits" : {
        "rate" : {
            "values": [
                {
                    "uri" : "/v1.1/*",
                    "regex" : "^/1.1/.*",
                    "limit" : [
                        {
                            "verb" : "GET",
                            "value" : 600000,
                            "remaining" : 426852,
                            "unit" : "HOUR",
                            "next-available" : "2011-02-22T19:32:43.835Z"
                        }
                    ]
                }
            ]
        },
        "absolute" : {
            "values" : {
                "maxNodesPerLoadBalancer" : 5,
                "maxVIPsperLoadBalancer" : 2
            }
        }
    }
}
```

3.8. Faults

API calls that return an error return one of the following fault objects. All fault objects extend from the base fault, `serviceFault`, for easier exception handling for languages that support it.

3.8.1. serviceFault

The `serviceFault` and by extension all other faults include `message` and `detail` elements which contain strings describing the nature of the fault as well as a `code` attribute representing the HTTP response code for convenience. The `code` attribute of the fault is for the convenience of the caller so that they may retrieve the response code from the HTTP response headers or directly from the fault object if they choose. The caller should not expect the `serviceFault` to be returned directly but should instead expect only one of the child faults to be returned.

3.8.2. badRequest

This fault indicates that the data in the request object is invalid; for example, a string was used in a parameter that was expecting an integer. The fault will wrap validation errors.

Example 3.3. Fault Response, badRequest

```
          <badRequest xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
 code="400">
              <message>Validation fault</message>
              <details>The object is not valid</details>
                  <validationErrors>
                      <message>Node ip is invalid. Please specify a valid ip.</
message>
                  </validationErrors>
          </badRequest>
```

3.8.3. immutableEntity

This fault is returned when a user attempts to modify an item that is not currently in a state that allows modification. For example, load balancers in a status of `PENDING_UPDATE`,`BUILD`, or `DELETED` may not be modified.

Example 3.4. Fault Response, immutableEntity

```
          <immutableEntity code="422" xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
              <message>The object at the specified URI is immutable and can not
 be overwritten.</message>
          </immutableEntity>
```

3.8.4. itemNotFound

Example 3.5. Fault Response, itemNotFound

```
          <itemNotFound code="404" xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
              <message>Object not Found</message>
          </itemNotFound>
```

3.8.5. loadBalancerFault

The `loadBalancerFault` fault shall be returned in the event that an error occurred during a loadbalancer operation.

Example 3.6. Fault Response, loadBalancerFault

```
          <loadBalancerFault code="500" xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
              <message>Load Balancer has experienced an internal error</message>
          </loadBalancerFault>
```

3.8.6. outOfVirtualIps

This fault indicates that there are no virtual IPs left to assign to a new load balancer. In practice, this fault should not occur, as virtual IPs will be ordered as capacity is required. If you do experience this fault, contact support so that we may make more IPs available.

Example 3.7. Fault Response, outOfVirtualIps

```
          <outOfVirtualIps code="500" xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
              <message>
                  Out of virtual IPs. Please contact support so they can allocate
 more virtual IPs.
              </message>
          </outOfVirtualips>
```

3.8.7. overLimit

This fault is returned when the user has exceeded a currently allocated limit.

Example 3.8. Fault Response, overLimit

```
          <overLimit code="413" xmlns="http://docs.openstack.org/loadbalancers/
api/v1.1">
              <message>Your account is currently over the limit so your request
 could not be processed.</message>
          </overLimit>
```

3.8.8. serviceUnavailable

This fault is returned when the service is unavailable, such as when the service is undergoing maintenance. Note that this does not necessarily mean that the currently configured loadbalancers are unable to process traffic; it simply means that the API is currently unable to service requests.

Example 3.9. Fault Response, serviceUnavailable

```
          <serviceUnavailable code="500" xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
              <message>The Load balancing service is currently not available</
message>
          </serviceUnavailable>
```

3.8.9. unauthorized

This fault is returned when the user is not authorized to perform an attempted operation.

Example 3.10. Fault Response, unauthorized

```
           <unauthorized code="401" xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
               <message>You are not authorized to execute this operation.</
message>
           </unauthorized>
```

3.8.10. unprocessableEntity

This fault is returned when an operation is requested on an item that does not support the operation - reword.

Example 3.11. Fault Response, unprocessableEntity

```
           <unprocessableEntity code="422" xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
               <message>The Object at the specified URI is unprocessable.</
message>
           </unprocessableEntity>
```

Chapter 4. API Operations

This chapter explains specific API operations. For ideas relevant to all API operations, see the "General API Information" chapter.

4.1. Load Balancers

4.1.1. List Load Balancers

Verb	URI	Description	Representation
GET	/loadbalancers	List all load balancers configured for the account.	XML, JSON

Normal Response Code(s): 200

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation provides a list of all load balancers configured and associated with your account.

To view deleted load balancers, add "?status=DELETED" to the end of the get url. A deleted load balancer is immutable and irrecoverable. Only a limited set of attributes will be returned in the response object:

- `id`
- `name`
- `algorithm`
- `protocol`
- `port`
- `created`
- `updated`

This operation does not require a request body.

Example 4.1. List Load Balancers Response: XML

```
<?xml version="1.0" ?>
<loadBalancers xmlns="http://docs.openstack.org/loadbalancers/api/v1.1">
    <loadBalancer id="71" name="lb-site1" status="ACTIVE"
        protocol="HTTP" port="80" algorithm="RANDOM">
        <virtualIps>
            <virtualIp id="403" address="206.55.130.1" ipVersion="IPV4"
                type="PUBLIC" />
        </virtualIps>
        <created time="2010-12-13T15:38:27-06:00" />
        <updated time="2010-12-13T15:38:38-06:00" />
    </loadBalancer>
    <loadBalancer id="166" name="lb-site2" status="ACTIVE"
        protocol="HTTP" port="80" algorithm="RANDOM">
        <virtualIps>
            <virtualIp id="401" address="206.55.130.2" ipVersion="IPV4"
                type="PUBLIC" />
        </virtualIps>
        <created time="2010-12-13T15:38:27-06:00" />
        <updated time="2010-12-13T15:38:38-06:00" />
    </loadBalancer>
</loadBalancers>
```

Example 4.2. List Load Balancers Response: JSON

```
{
    "loadBalancers":[
        {
            "name":"lb-site1",
            "id":"71",
            "protocol":"HTTP",
            "port":"80",
            "algorithm":"RANDOM",
            "status":"ACTIVE",
            "virtualIps":[
                {
                    "id":"403",
                    "address":"206.55.130.1",
                    "type":"PUBLIC",
                    "ipVersion":"IPV4"
                }
            ],
            "created":{
                "time":"2010-11-30T03:23:42Z"
            },
            "updated":{
                "time":"2010-11-30T03:23:44Z"
            }
        },
        {
            "name":"lb-site2",
            "id":"166",
            "protocol":"HTTP",
            "port":"80",
            "algorithm":"RANDOM",
            "status":"ACTIVE",
            "virtualIps":[
                {
                    "id":"401",
                    "address":"206.55.130.2",
                    "type":"PUBLIC",
                    "ipVersion":"IPV4"
                }
            ],
            "created":{
                "time":"2010-11-30T03:23:42Z"
            },
            "updated":{
                "time":"2010-11-30T03:23:44Z"
            }
        }
    ]
}
```

4.1.2. List Load Balancer Details

Verb	URI	Description	Representations
GET	/loadbalancers/loadBalancerId	List details of the specified load balancer.	JSON, XML

Normal Response Code(s): 200

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation provides detailed output for a specific load balancer configured and associated with your account. This operation is not capable of returning details for a load balancer which has been deleted.

This operation does not require a request body.

Example 4.3. List Load Balancer Details Request: XML

```
<loadBalancer xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    id="2000"
    name="sample-loadbalancer"
    protocol="HTTP"
    port="80"
    algorithm="ROUND_ROBIN"
    status="ACTIVE">
    <connectionLogging enabled="false" />
    <virtualIps>
        <virtualIp
            id="1000"
            address="206.10.10.210"
            type="PUBLIC"
            ipVersion="IPV4" />
    </virtualIps>
    <nodes>
        <node
            nodeId="1041"
            address="10.1.1.1"
            port="80"
            condition="ENABLED"
            status="ONLINE" />
        <node
            nodeId="1411"
            address="10.1.1.2"
            port="80"
            condition="ENABLED"
            status="ONLINE" />
    </nodes>
    <sessionPersistence persistenceType="HTTP_COOKIE"/>
    <connectionThrottle
        maxConnectionRate="50"
        rateInterval="60" />
    <cluster name="c1.dfw1" />
    <created time="2010-11-30T03:23:42Z" />
    <updated time="2010-11-30T03:23:44Z" />
</loadBalancer>
```

Example 4.4. List Load Balancers Details Response: JSON


```
{
    "loadBalancer":{
        "id": 2000,
        "name":"sample-loadbalancer",
        "protocol":"HTTP",
        "port": 80,
        "algorithm":"ROUND_ROBIN",
        "status":"ACTIVE",
        "connectionLogging":{
            "enabled":"true"
        },
        "virtualIps":[
            {
                "id": 1000,
                "address":"206.10.10.210",
                "type":"PUBLIC",
                "ipVersion":"IPV4"
            }
        ],
        "nodes":[
            {
                "id": 1041,
                "address":"10.1.1.1",
                "port": 80,
                "condition":"ENABLED",
                "status":"ONLINE"
            },
            {
                "id": 1411,
                "address":"10.1.1.2",
                "port": 80,
                "condition":"ENABLED",
                "status":"ONLINE"
            }
        ],
        "sessionPersistence":{
            "persistenceType":"HTTP_COOKIE"
        },
        "connectionThrottle":{
            "maxRequestRate": 50,
            "rateInterval": 60
        },
        "cluster":{
            "name":"c1.dfw1"
        },
        "created":{
            "time":"2010-11-30T03:23:42Z"
        },
        "updated":{
            "time":"2010-11-30T03:23:44Z"
        }
    }
}
```

4.1.3. Create Load Balancer

Verb	URI	Description	Representation
POST	/loadbalancers	Create a new load balancer with the configuration defined by the request.	XML, JSON

Normal Response Code(s): 202

Error Response Code(s): loadbalancerFault (500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation provisions a new load balancer based on the configuration defined in the request object. Once the request is validated and progress has started on the provisioning process, a response object will be returned. The object will contain a unique identifier and status of the request.

If the status returned is set to "BUILD Using the identifier, the caller can check on the progress of the operation by performing a **GET** on `loadbalancers/id`. When the status of the loadBalancer changes to "ACTIVE", then the loadbalancer has been successfully provisioned and is operational.

The caller of this operation must specify at least the following attributes of the loadbalancer:

- `name`
- `protocol`
- `port`
- `At least one VirtualIp`
- `At least one node`

If the request cannot be fulfilled due to insufficient or invalid data, an `HTTP` 400 (Bad Request) error response will be returned with information regarding the nature of the failure in the body of the response. Failures in the validation process are non-recoverable and require the caller to correct the cause of the failure and **POST** the request again.

Note

A load balancer's name has a max length that can be queried when querying limits.

Note

Users may configure all documented features of the load balancer at creation time by simply providing the additional elements or attributes in the request. This document provides an overview of all the features the load balancing service supports.

Note

Users may request either a IPv4 or IPV6 address by specifying the version in the create request. To request a IPV6 virtual address the version is specified as

```
<virtualIp type="PUBLIC" version="IPV6"/>
```

Example 4.5. Create Load Balancer (Required Attributes) Request: XML

```
<loadBalancer xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    name="a-new-loadbalancer"
    port="80"
    protocol="HTTP">
    <virtualIps>
        <virtualIp type="PUBLIC"/>
    </virtualIps>
    <nodes>
        <node address="10.1.1.1" port="80" condition="ENABLED"/>
    </nodes>
</loadBalancer>
```

Example 4.6. Create Load Balancer (Required Attributes) Request: JSON

```
{
    "loadBalancer": {
        "name": "a-new-loadbalancer",
        "port": "80",
        "protocol": "HTTP",
        "virtualIps": [
            {
                "type": "PUBLIC"
            }
        ],
        "nodes": [
            {
                "address": "10.1.1.1",
                "port": "80",
                "condition": "ENABLED"
            }
        ]
    }
}
```

If you have at least one load balancer, you may create subsequent load balancers that share a single virtual IP by issuing a **POST** and supplying a virtual IP ID instead of a type. Additionally, this feature is highly desirable if you wish to load balance both an unsecured and secure protocol using one IP or DNS name. For example, this method makes it possible to use the same load balancing configuration to support HTTP and HTTPS).

Note

Load balancers sharing a virtual IP *must* utilize a unique port.

Example 4.7. Create Load Balancer (Required Attributes with Shared IP) Request: XML

```
<loadBalancer xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    name="a-new-loadbalancer"
    port="80"
    protocol="HTTP">
    <virtualIps>
        <virtualIp id="39"/>
    </virtualIps>
    <nodes>
        <node address="10.1.1.1" port="80" condition="ENABLED" />
    </nodes>
</loadBalancer>
```

Example 4.8. Create Load Balancer (Required Attributes with Shared IP) Request: JSON

```
{
    "loadBalancer":{
        "name":"a-new-loadbalancer",
        "port":"80",
        "protocol":"HTTP",
        "virtualIps":[
            {
                "id":"39"
            }
        ],
        "nodes":[
            {
                "address":"10.1.1.1",
                "port":"80",
                "condition":"ENABLED"
            }
        ]
    }
}
```

Example 4.9. Create Load Balancer (Required Attributes with Shared IP) Response: XML

```
<loadBalancer xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    id="144"
    name="a-new-loadbalancer"
    algorithm="ROUND_ROBIN"
    protocol="HTTP"
    port="83"
    status="BUILD">
    <virtualIps>
        <virtualIp
            id="39"
            address="206.10.10.210"
            ipVersion="IPV4"
            type="PUBLIC" />
    </virtualIps>
    <nodes>
        <node
            id="653"
            address="10.1.1.1"
            port="80"
            condition="ENABLED"
            status="ONLINE" />
    </nodes>
    <cluster name="ztm-n03.staging1.lbaas.demo.net" />
    <created time="2011-02-08T21:19:55Z" />
    <updated time="2011-02-08T21:19:55Z" />
    <connectionLogging enabled="false" />
</loadBalancer>
```

Example 4.10. Create Load Balancer (Required Attributes with Shared IP) Response: JSON

```
{
    "loadBalancer": {
        "name": "a-new-loadbalancer",
        "id": 144,
        "protocol": "HTTP",
        "port": 83,
        "algorithm": "ROUND_ROBIN",
        "status": "BUILD",
        "cluster": {
            "name": "ztm-n01.staging1.lbaas.demo.net"
        },
        "nodes": [
            {
                "address": "10.1.1.1",
                "id": 653,
                "port": 80,
                "status": "ONLINE",
                "condition": "ENABLED"
            }
        ],
        "virtualIps": [
            {
                "address": "206.10.10.210",
                "id": 39,
                "type": "PUBLIC",
                "ipVersion": "IPV4"
            }
        ],
        "created": {
            "time": "2011-04-13T14:18:07Z"
        },
        "updated": {
            "time": "2011-04-13T14:18:07Z"
        },
        "connectionLogging": {
            "enabled": false
        }
    }
}
```

4.1.4. Update Load Balancer Attributes

Verb	URI	Description	Representation
PUT	/loadbalancers/*loadBalancerId*	Update the properties of a load balancer.	XML, JSON

Normal Response Code(s): 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation updates the attributes of the specified load balancer. Upon successful validation of the request, the service will return a 202 (Accepted) response code. A caller should check that the load balancer status is ACTIVE to confirm that the update has taken effect. If the load balancer status is

"PENDING_UPDATE" then the caller can poll the load balancer with its ID (using a GET operation) to wait for the changes to be applied and the load balancer to return to an ACTIVE status.

This operation allows the caller to change one or more of the following attributes:

- name
- algorithm

This operation does not return a response body.

Note

The load balancer's ID, status, port and protocol are immutable attributes and cannot be modified by the caller. Supplying an unsupported attribute will result in a 400 (badRequest) fault.

Example 4.11. Update Load Balancer Attributes Request: XML

```
<loadBalancer xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    name="sample-loadbalancer"
    algorithm="ROUND_ROBIN" />
```

Example 4.12. Update Load Balancer Attributes Request: JSON

```
{"loadBalancer":{
    "name": "sample-loadbalancer",
    "algorithm": "ROUND_ROBIN"
    }
}
```

Theload balancer's tatus attribute reflects the current configuration status of the device. This status is immutable by the caller and is updated automatically based on state changes within the service. When a load balancer is first created, it may be placed into a BUILD status while the configuration is being generated and applied based on the request. Once the configuration is applied and finalized, it will be in an ACTIVE status. In the event of a configuration change or update, the status of the load balancer may change to PENDING_UPDATE to signify configuration changes are in progress but have not yet been finalized. Load balancers in a SUSPENDED status reject traffic and will not forward requests to back-end nodes.

Table 4.1. Load Balancer Statuses

Name	Description
ACTIVE	Load balancer is configured properly and ready to serve traffic to incoming requests via the configured virtual IPs.
BUILD	Load balancer is being provisioned for the first time and configuration is being applied to bring the service online. The service will not yet be ready to serve incoming requests.
PENDING_UPDATE	Load balancer is online but configuration changes are being applied to update the service based on a previous request.
PENDING_DELETE	Load balancer is online but configuration changes are being applied to begin deletion of the service based on a previous request.
SUSPENDED	Load balancer has been taken offline and disabled.
ERROR	The system encountered an error when attempting to configure the load balancer.

Name	Description
DELETED	Load balancers in DELETED status can be displayed for a certain number of days after deletion. The number of days is queryable.

4.1.5. Remove Load Balancer

Verb	URI	Description
DELETE	/loadbalancers/*loadBalancerId*	Remove a load balancer from the account.

Normal Response Code(s): 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

The remove load balancer function removes the specified load balancer and its associated configuration from the account. Any and all configuration data is immediately purged and is not recoverable.

This operation does not require a request body.

This operation does not return a response body.

4.2. Nodes

The nodes defined by the load balancer are responsible for servicing the requests received through the load balancer's virtual IP. By default, the load balancer employs a basic health check that ensures the node is listening on its defined port. The node is checked at the time of addition and at regular intervals as defined by the load balancer health check configuration. If a back-end node is not listening on its port or does not meet the conditions of the defined active health check for the load balancer, then the load balancer will not forward connections and its status will be listed as OFFLINE. Only nodes that are in an ONLINE status will receive and be able to service traffic from the load balancer.

The status of the node can be determined by passive or active health monitoring.

The caller can assign the relevant weights to the node using the weight attribute of the node element. The weight of a node determines the portion of requests it services compared to the other nodes of the load balancer. For exampl, if node A has a weight of 2 and node B has a weight of 1, then the load balancer will forward twice as many requests to node A than to node B. If the weight attribure is not specified, then the node's weight is implicitly set to "1".

4.2.1. List Nodes

Verb	URI	Description
GET	/loadbalancers/*loadBalancerId*/nodes	List node(s) configured for the load balancer.
GET	/loadbalancers/*loadBalancerId*/nodes/*nodeId*	List details of a specific node.

Normal Response Code(s): 200

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

Example 4.13. List Node Response: XML

```
<node
    id="410"
    address="10.1.1.1"
    port="80"
    condition="ENABLED"
    status="ONLINE" />
```

Example 4.14. List Node Response: JSON

```
{"node": {
        "id":"410",
        "address":"10.1.1.1",
        "port":80,
        "condition":"ENABLED",
        "status":"ONLINE"
    }
}
```

Example 4.15. List Nodes Response: XML

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<nodes xmlns="http://docs.openstack.org/loadbalancers/api/v1.1">
    <node
            id="650"
            address="10.1.1.1"
            port="80"
            condition="ENABLED"
            status="ONLINE"/>
    <node
            id="183"
            address="10.2.2.1"
            port="80"
            condition="ENABLED"
            status="ONLINE"/>
    <node
            id="184"
            address="10.2.2.2"
            port="88"
            condition="ENABLED"
            status="ONLINE"/>
</nodes>
```

Example 4.16. List Nodes Response: JSON

```
{
    "nodes": [
        {
            "address": "10.1.1.1",
            "id": 650,
            "port": 80,
            "status": "ONLINE",
            "condition": "ENABLED"
        },
        {
            "address": "10.2.2.1",
            "id": 183,
            "port": 80,
            "status": "ONLINE",
            "condition": "ENABLED"
        },
        {
            "address": "10.2.2.2",
            "id": 184,
            "port": 88,
            "status": "ONLINE",
            "condition": "ENABLED"
        }
    ]
}
```

4.2.2. Add Nodes

Verb	URI	Description
POST	/loadbalancers/*loadBalancerId*/nodes	Add a new node to the load balancer.

Normal Response Code(s): 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

When a node is added, it is assigned a unique identifier that can be used for mutating operations such as changing the condition or removing it. Every load balancer is dual-homed on both the public Internet and internal; as a result, nodes can either be internal private addresses or addresses on the public Internet.

Example 4.17. Add Nodes Request: XML

```
<nodes xmlns="http://docs.openstack.org/loadbalancers/api/v1.1">
    <node address="10.1.1.1" port="80" condition="ENABLED" />
    <node address="10.2.2.1" port="80" condition="ENABLED" />
    <node address="10.2.2.2" port="88" condition="ENABLED" />
</nodes>
```

Example 4.18. Add Nodes Request: JSON

```
{"nodes": [
        {
            "address": "10.1.1.1",
            "port": 80,
            "condition": "ENABLED"
        },
        {
            "address": "10.2.2.1",
            "port": 80,
            "condition": "ENABLED"
        },
        {
            "address": "10.2.2.2",
            "port": 88,
            "condition": "ENABLED",
            "weight": 10
        }
    ]
}
```

Example 4.19. Add Nodes Response: XML

```
<nodes xmlns="http://docs.openstack.org/loadbalancers/api/v1.1">
    <node
        address="10.2.2.3"
        id="185"
        port="80"
        condition="ENABLED"
        status="ONLINE" />
    <node
        address="10.2.2.4"
        id="186"
        port="80"
        condition="ENABLED"
        status="ONLINE" />
</nodes>
```

Example 4.20. Add Nodes Response: JSON

```
{"nodes": [
        {
            "address": "10.2.2.3",
            "id": 185,
            "port": 80,
            "status": "ONLINE",
            "condition": "ENABLED"
        },
        {
            "address": "10.2.2.4",
            "id": 186,
            "port": 80,
            "status": "ONLINE",
            "condition": "ENABLED"
        }
    ]
}
```

4.2.3. Modify Nodes

Verb	URI	Description
PUT	/loadbalancers/*loadBalancerId*/nodes/*nodeId*	Modify the configuration of a node on the load balancer.

Normal Response Code(s): 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

Note

The node's IP and port are immutable attributes and cannot be modified with a **PUT** request. Supplying an unsupported attribute will result in a 400 (badRequest) fault. A load balancer supports a maximum number of nodes. The maximum number of nodes per loadbalancer is a queryable limit.

Every node in the load balancer has an associated condition which determines its role within the load balancer.

Table 4.2. Load Balancer Node Conditions

Name	Description
`ENABLED`	Node is permitted to accept new connections. Its status will eventually become ONLINE to reflect this configuration.
`DISABLED`	Node is not permitted to accept any new connections regardless of session persistence configuration. Existing connections are forcibly terminated. The node's status will change to OFFLINE once the configuration has been successfully appplied.

Example 4.21. Update Node Condition Request: XML

```
<node condition="ENABLED" />
```

Example 4.22. Update Node Condition Request: JSON

```
{"node":{
        "condition": "ENABLED"
    }
}
```

4.2.4. Remove Nodes

Verb	URI	Description
DELETE	/loadbalancers/*loadBalancerId*/nodes/*nodeId*	Remove a node from the load balancer.

Normal Response Code(s): 200, 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

4.3. Virtual IPs

A virtual IP (VIP) makes a load balancer accessible by clients. The load balancing service supports either a public VIP, routable on the public Internet, or a private address, routable only within the region in which the load balancer resides.

Table 4.3. Virtual IP Types

Name	Description
PUBLIC	An address that is routable on the public Internet.
INTERNAL	An address that is routable only on internal network.

4.3.1. List Virtual IPs

Verb	URI	Description
GET	/loadbalancers/*loadBalancerId*/virtualips	List all virtual IPs associated with a load balancer.

Normal Response Code(s): 200, 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This request does not require a request body.

Note

The number of VIPs that can be configured when creating a load balancer is a queryable limit.

Example 4.23. List Virtual IPs Response: XML

```
                        <virtualIps xmlns="http://docs.openstack.org/
loadbalancers/api/v1.1">
    <virtualIp
        id="1000"
        address="206.10.10.210"
        type="PUBLIC"/>
</virtualIps>
```

Example 4.24. List Virtual IPs Response: JSON

```
                        {"virtualIps": [
    {
    "id": "1000",
    "address": "206.10.10.210",
    "type": "PUBLIC"
    }
  ]
}
```

4.4. Usage Reports

4.4.1. List Usage

Name	URI	Description
GET	/loadbalancers/loadBalancerId/usage	List current and historical usage.

Normal Response Code(s): 200

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

The load balancer usage reports provide a set of usage counters. This list will contain at least the transferBytesIn and transferBytesOut usage counters that represent respectively the amount of traffic in bytes received by the load balancer from clients requests, and the amount fo traffic sent from the load balancer as responses to clients.

Example 4.25. Report Load Balancer Usage Response: XML

```
<loadBalancerUsage xmlns="http://docs.openstack.org/loadbalancers/api/v1.1">
    <loadBalancerUsageRecord
            id="394"
            transferBytesIn="0"
            transferBytesOut="0" />
    <loadBalancerUsageRecord
            id="473"
            transferBytesIn="0"
            transferBytesOut="0" />
</loadBalancerUsage>
```

Example 4.26. Report Load Balancer Usage Response: JSON

```
{
    "loadBalancerUsageRecords": [
        {
            "id": "394",
            "transferBytesIn": "0",
            "transferBytesOut": "0"
        },
        {
            "id": "473",
            "transferBytesIn": "0",
            "transferBytesOut": "0"
        }
    ]
}
```

4.5. Monitors

In addition to the default passive monitoring, the load balancing service includes an active health monitoring operation which periodically checks your back-end nodes to ensure they are responding correctly.

Active health monitoring provides 3 types of health monitors. The caller can configure one health monitor on the load blancer.

The health monitor has a `type` attribute to signify what kind of monitor it is. This specification supports 3 types of health monitor.

Table 4.4. Health Monitor Types

Name	Description
CONNECT	Health monitor is a CONNECT monitor.
HTTP	Health monitor is an HTTP monitor.
HTTPS	Health monitor is an HTTPS monitor.

4.5.1. Monitor Health

Verb	URI	Description
GET	/loadbalancers/*loadBalancerId*/healthmonitor	Retrieve the health monitor configuration, if one exists.
PUT	/loadbalancers/*loadBalancerId*/healthmonitor	Update the settings for a health monitor.
DELETE	/loadbalancers/*loadBalancerId*/healthmonitor	Remove the health monitor.

Normal Response Code(s): 200, 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

4.5.1.1. Monitor CONNECT

The monitor connects to each node on its defined port to ensure that the service is listening properly. The CONNECT monitor is the most basic type of health check and does no post-processing or protocol specific health checks. It includes several configurable properties:

- `delay`: This is the minimum time between calls to a monitor.
- `timeout`: Maximum number of seconds to wait for a connection to be established before timing out. The value must be less than the delay value.
- `attemptsBeforeDeactivation`: Number of permissible monitor failures before removing a node from rotation. (Must be a number between 1 and 10)

Example 4.27. Monitor Connections Request: XML

```
<healthMonitor xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    type="CONNECT"
    delay="20"
    timeout="10"
    attemptsBeforeDeactivation="3" />
```

Example 4.28. Monitor Connections Request: JSON

```
{"healthMonitor": {
        "type": "CONNECT",
        "delay": 20,
        "timeout": 10,
        "attemptsBeforeDeactivation": 3
   }
}
```

Example 4.29. Monitor Connections Response: XML

```
<healthMonitor xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    type="CONNECT"
    delay="10"
    timeout="10"
    attemptsBeforeDeactivation="3" />
```

Example 4.30. Monitor Connections Response: JSON

```
{"healthMonitor": {
        "type": "CONNECT",
        "delay": 20,
        "timeout": 10,
        "attemptsBeforeDeactivation": 3
   }
}
```

4.5.1.2. Monitor HTTP and HTTPS

The HTTP and HTTPS monitor is more intelligent than the connect monitor. It is capable of processing an HTTP or HTTPS response to determine the condition of a node. It supports the same basic properties as the CONNECT monitor and includes additional attributes of path that is used to evaluate the HTTP response.

- `path`: The HTTP path used in the request by the monitor. This must be a string beginning with a / (forward slash). The monitor expects a response from teh node with an HTTP status code of 200.

Example 4.31. Monitor HTTP Response: XML

```
<healthMonitor xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    type="HTTP"
    delay="10"
    timeout="3"
    attemptsBeforeDeactivation="2"
    path="/"
/>
```

Example 4.32. Monitor HTTPS Response: XML

```
<healthMonitor xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    type="HTTPS"
    delay="10"
    timeout="3"
    attemptsBeforeDeactivation="2"
    path="/"
 />
```

4.6. Sessions

4.6.1. Manage Session Persistence

Session persistence is a feature of the load balancing service which forces multiple requests from clients to be directed to the same node. This is common with many web applications that do not inherently share application state between back-end servers.

Table 4.5. Session Persistence Modes

Name	Description
HTTP_COOKIE	A session persistence mechanism that inserts an HTTP cookie and is used to determine the destination back-end node. This is supported for HTTP load balancing only.

Verb	URI	Description
GET	/loadbalancers/loadBalancerId/sessionpersistence	List session persistence configuration.
PUT	/loadbalancers/loadBalancerId/sessionpersistence	Enable session persistence.
DELETE	/loadbalancers/loadBalancerId/sessionpersistence	Disable session persistence.

Normal Response Code(s): 200, 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

Example 4.33. List Session Persistence Configuration Response: XML

```
<sessionPersistence xmlns="http://docs.openstack.org/loadbalancers/api/v1.
1" persistenceType="HTTP_COOKIE"/>
```

Example 4.34. List Session Persistence Configuration Response: JSON

```
{
   "sessionPersistence":{
      "persistenceType":"HTTP_COOKIE"
   }
}
```

Example 4.35. Set Session Persistence Type Request: XML

```
<sessionPersistence xmlns="http://docs.openstack.org/loadbalancers/api/v1.
1" persistenceType="HTTP_COOKIE"/>
```

Example 4.36. Set Session Persistence Type Request: JSON

```
{
   "sessionPersistence":{
      "persistenceType":"HTTP_COOKIE"
   }
}
```

4.7. Connections

4.7.1. Log Connections

Verb	URI	Description
GET	/loadbalancers/loadBalancerId/connectionlogging	View current configuration of connection logging.
PUT	/loadbalancers/loadBalancerId/connectionlogging	Enable or disable connection logging.

Normal Response Code(s): 200, 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation allows the user to view the current connection logging configuration, enable connection logging, or disable connection logging.

This operation does not require a request body.

The service provider is responsible for providing an out-of-brand method for users of the service to retrieve the access and connection logs of their load balancers.

Example 4.37. List Connection Logging Configuration Response: XML

```
<connectionLogging xmlns="http://docs.openstack.org/loadbalancers/api/v1.
1" enabled="true"/>
```

Example 4.38. List Connection Logging Configuration Response: JSON

```
{
   "connectionLogging": {
      "enabled": "true"
   }
}
```

Example 4.39. Enable Connection Logging Request: XML

```
<connectionLogging xmlns="http://docs.openstack.org/loadbalancers/api/v1.
1" enabled="true"/>
```

Example 4.40. Enable Connection Logging Request: JSON

```
{
   "connectionLogging":{
      "enabled":"true"
   }
}
```

4.7.2. Throttle Connections

Verb	URI	Description
GET	/loadbalancers/loadBalancerId/connectionthrottle	List connection throttling configuration.
PUT	/loadbalancers/loadBalancerId/connectionthrottle	Update throttling configuration.
DELETE	/loadbalancers/loadBalancerId/connectionthrottle	Remove connection throttling configurations.

Normal Response Code(s): 200, 202

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

The connection throttling feature imposes limits on the number of connections per IP address to help mitigate malicious or abusive traffic to your applications. The following properties can be configured based on the traffic patterns for your sites.

Note

Whent the rate is exceeded, the load balancer returns a serviceUnavailable (503) for HTTP/HTTPS loadbalancers and refuses connections for TCP loadbalancers.

- `maxRequestRate`: Maximum number of requests allowed from a single IP address in the defined `rateInterval`. Setting a value of 0 allows an unlimited connection rate.
- `rateInterval`: Frequency (in seconds) at which the `maxRequestRate` is assessed. For example, a `maxRequestRate` of 30 with a `rateInterval` of 60 would allow a maximum of 30 connections per minute for a single IP address. This value must be specified between 1 and 3600.

Example 4.41. List Connection Throttling Configuration Response: XML

```
<connectionThrottle xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    maxRequestRate="50"
    rateInterval="60" />
```

Example 4.42. List Connection Throttling Configuration Response: JSON

```
{"connectionThrottle":{
    "maxRequestRate": 50,
    "rateInterval": 60
  }
}
```

Example 4.43. Update Connection Throttling Configuration Request: XML

```
<connectionThrottle xmlns="http://docs.openstack.org/loadbalancers/api/v1.1"
    maxRequestRate="50"
    rateInterval="60" />
```

Example 4.44. Update Connection Throttling Configuration Request: JSON

```
{"connectionThrottle":{
    "maxRequestRate": 50,
    "rateInterval": 60
 }
}
```

4.8. Protocols

4.8.1. List Load Balancing Protocols

Verb	URI	Description
GET	/loadbalancers/protocols	List all supported load balancing protocols.

Normal Response Code(s): 200

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

All load balancers must define the protocol of the service which is being load balanced. The protocol selection should be based on the protocol of the back-end nodes. When configuring a HTTP or HTTPS load balancer, the default port for the given protocol will be selected unless otherwise specified. For TCP load balancers, the prot attribute must be provided.

Example 4.45. List Load Balancing Protocols Response: XML

```
<protocols xmlns="http://docs.openstack.org/loadbalancers/api/v1.1">
    <protocol name="TCP"/>
    <protocol name="HTTP" port="80" />
    <protocol name="HTTPS" port="443" />
</protocols>
```

Example 4.46. List Load Balancing Protocols Response: JSON

```
{"protocols": [
        {
            "name": "HTTP",
            "port": "80"
        },
        {
            "name": "HTTPS",
            "port": "443"
        },
        {
            "name": "TCP"
        }
    ]
}
```

4.9. Algorithms

All load balancers utilize an algorithm that defines how traffic should be directed between back-end nodes. The default algorithm for newly created load balancers is `ROUND-ROBIN`, which can be overridden at creation time or changed after the load balancer has been initially provisioned. The algorithm name is to be constant within a major revision of the load balancing API, though new algorithms may be created with a unique algorithm name within a given major revision of the service API.

Table 4.6. Load Balancing Algorithms

Name	Description
`LEAST_CONNECTIONS`	The node with the lowest number of connections will receive requests. Weights can be defined as part of the node configuration.
`ROUND_ROBIN`	Connections are routed to each of the back-end servers in turn. This is the default algorithm. Weights can be defined as part of the node configuration.

4.9.1. List Load Balancing Algorithms

Verb	URI	Description
GET	/loadbalancers/algorithms	List all supported load balancing algorithms.

Normal Response Code(s): 200

Error Response Code(s): loadbalancerFault (400, 500), serviceUnavailable (503), unauthorized (401), badRequest (400), overLimit (413)

This operation does not require a request body.

Example 4.47. List Load Balancing Algorithms Response: XML

```
<algorithms xmlns="http://docs.openstack.org/loadbalancers/api/v1.1">
    <algorithm name="LEAST_CONNECTIONS" />
    <algorithm name="ROUND_ROBIN" />
</algorithms>
```

Example 4.48. List Load Balancing Algorithms Response: JSON

```
{"algorithms": [
        {
            "name": "LEAST_CONNECTIONS"
        },
        {
            "name": "ROUND_ROBIN"
        }
    ]
}
```

Chapter 5. API Extensions

Implementations of this API specifications are free to augment this core API with extensions as they see appropriate to extend Load Balancing features (e.g. support for new LB algorithms)or offer new ones. All client applications written to this core specification however (using this core API) must be supported by extended implementations as specified in this document. Therefore, these applications should not receive payloads or values not specified in this specification or obtain a different behavior from the service than expected.

For a detailed description of how to use Extension APIs in OpenStack services, refer to OpenStack Compute API1.1 documentation on the OpenStack website.