Enable-U
The Integration Specialist

Kaap Noord	+31 (0)20 716 38 66	CC	nl-34228397
Asterweg 19D12	www.enable-u.nl	VAT	nl8147.71.725.BO1
1031 HL Amsterdam	info@enable-u.nl	IBAN	NL67RABO0114665

Service Description: Load Balancer Status Service

Summary
Service name: Load Balancer Status Service
Version: 1.0
Date: 02 Nov 2017
Short description: Sends back a status response configured for each node in the cluster so the load balancer behavior can be controlled from the gateway.

Use Case
To meet availability and/or performance requirements, the CA API Gateway can be implemented in a cluster with multiple gateway nodes. As part of the cluster setup, an external load balancer must be configured to direct traffic to the different nodes in the cluster.

In general, a load balancer will check each node for health in combination with a manual configuration to decide if it should send traffic to a node. Most load balancers recognize 3 statuses: healthy, drain and unavailable, which can be either manually configured by the load balancer administrator or automatically triggered by the load balancer based on node status. A healthy node can receive traffic, drain means that existing session can be completed but no new sessions should be sent to a node while an unavailable node should not receive any traffic from the load balancer.

Because the load balancer is usually not maintained by the API Gateway administrators, it will require involvement of an additional person, the load balancer administrator, when doing maintenance with downtime on the cluster nodes. However, if a load balancer supports determining node status based on the content of a http response from each node, a service can be used to control the status on the load balancer. This way an API Gateway administrator can do maintenance on cluster nodes without the need to involve the load balancer administrator.

Description
This service can be used to provide a http response for a health check from a load balancer. By default, it will return a http 200 response code with a plain text message which can be either OK, DRAIN or OFFLINE, depending on the setting which was selected in the policy.

Kaap Noord +31 (0)20 716 38 66 cc nl-34228397
Asterweg 19D12 www.enable-u.nl VAT nl8147.71.725.BO1
1031 HL Amsterdam info@enable-u.nl IBAN NL67RABO0114665

Implementation

To implement the Load Balancer Status Service, you need to publish a new Web API in the Policy Manager. Description and resolution path can be filled in according to existing naming conventions, for example Load Balancer Status Service (v1.0) for the name and /lbstatus/v1 as path. An * is not required in the resolution path. Optionally the http method can be limited to GET requests for added security.

Once the service is created, the Load Balancer Status Service policy xml file can be imported in the policy manager. Additional nodes or a change in the response will require a change in the main part of the policy. It is not required to remove anything in the policy if less nodes are present than what the policy was created for. Node names and node status can be configured in the configuration section of the policy.

Configuration

The Load Balancer Status Service policy contains a configuration section to easily configure the most common changes in the policy. The Configuration section contains 3 main parts:

- Security: here you can add or change access restrictions so only the load balancer and other authorized systems, like monitoring systems, will be able to access the service. The default configuration is based on an IP whitelist for 192.168.0.0/12.
- Node names: here you configure the names of the cluster nodes. The node name can be found and set in the Policy Manager main menu option View -> Dashboard on the Cluster Status tab. The default node names are Gateway1 and Gateway2
- Node status: in this section you can select the active status for each node by simply enabling the right setting and disabling all other options. If multiple status settings are active for one node, the last one will be used.

Usage

Once you have implemented and configured the Load Balancer Status Service and activated it, it is ready to be used by the load balancer. Configure the load balancer to call the configured resolution path (for example /lbstatus/v1) on each of the cluster nodes and to automatically adjust node status according to the response it gets from each node. Once this is done, the API Gateway administrator should be able to control load balancer behavior without the need to involve the load balancer administrator, simply by adjusting the settings in the policy and activating the new settings in the API Gateway.

Release history

Version	Date	Author	Comments
1.0	02 Nov 2017	Michiel Helder	First release