vRealize Automation Load Balancing

Configuration Guide
Version 7.0.x

TECHNICAL WHITE PAPER

JUNE 2016

VERSION 2.3

Table of Contents

Revision History

DATE	VERSION	DESCRIPTION
August 2015	1.0	Initial version
December 2015	1.1	Minor updates
December 2015	2.0	Updates for vRealize Automation 7.0
January 2016	2.1	Minor updates
May 2016	2.2	▪ Updates for vRealize Automation 7.0.x
June 2016	2.3	▪ Updated timeout to 10 seconds for Configure Monitors and Add Service Monitoring in F5 and NSX sections respectively ▪ Added source IP persistence and timeout of 1800 seconds for Add Application Profiles section ▪ Updated all the screenshots to match the content ▪ Updated NSX load balancing method to be round robin

Introduction

This document describes the configuration of the load balancing modules of F5 Networks BIG-IP software (F5) and NSX load balancers for vRealize Automation 7.0.x in a distributed and high availability deployment. This document is not an installation guide, but a load-balancing configuration guide that supplements the vRealize Automation installation and configuration documentation available in the *vRealize Automation Installation and Configuration* guide in the VMware vRealize Automation 7.0 Documentation Center.

This information is for the following products and versions.

PRODUCT	VERSION
F5 BIG IP	Tested with 11.6
NSX	6.1.3, 6.1.4 (versions below 6.1.3 are not supported)
vRealize Automation	7.0.x

Load Balancing Concepts

Load balancers distribute work among servers in high-availability deployments. The system administrator backs up the load balancers on a regular basis at the same time as other components.

Follow your site policy for backing up load balancers, keeping in mind the preservation of network topology and vRealize Automation backup planning.

SSL Pass-Through

SSL pass-through is used with the load balancing configurations for the following reasons:

- **Ease of deployment**. Not having to deploy the vRealize Automation certificates to the load balancer simplifies deployment and reduces complexity.
- **No operational overhead**. At the time of certificate renewal, no configuration changes are required on the load balancer.
- **Ease of communication**. The individual host names of the load-balanced components are in the subject alternate name field of the certificates, so the client has no problem communicating with the load balanced nodes.

Session Persistence

The persistence option overrides any load balancing algorithm option, for example: setting dest_addr overrides, setting round robin, and so on. Different components in the vRealize Automation architecture benefit from different persistence methods. The configuration recommended in this document is the result of extensive testing and represents the best compromise between stability, performance, and scalability.

Destination Address (F5)

Destination address affinity persistence, also known as sticky persistence, supports TCP and UDP protocols, and directs session requests to the same server based on the destination IP address of a packet.

Source (IP) Address (F5 & NSX)

The default source IP address persistence option persists traffic based on the source IP address of the client for the life of that session and until the persistence entry timeout expires. The default for this persistence is 180 seconds. The next time a persistent session from that same client is initiated, it might be persisted to a different member of the pool. This decision is made by the load balancing algorithm and is non-deterministic.

NOTE: Set the persistence entry timeout to 1800 seconds (30 minutes) to match the vRealize Automation GUI

timeout.

Source IP Address Hash (NSX)

The source IP address is hashed and divided by the total weight of the running servers to designate which server receives the request. This process ensures that the same client IP address always reaches the same server if no server fails or starts. For more information on IP Hash load balancing, see VMware Knowledge base article KB 2006129.

Email notifications on Load Balancer

It is a good practice to set up an email notification on the Load Balancer that sends emails to the system administrator every time a vRA/vRO node goes down. Currently, NSX does not support email notification for such a scenario.

You can set up an email notification with F5 by following methods:

- Configuring the BIG-IP system to deliver locally generated email messages
- Configuring custom SNMP traps
- Configuring alerts to send email notifications

One-arm or Multiarm Topologies

In one-arm deployment, the components to be load balanced and the load balancers' virtual IP (VIP) are on the same network. Traffic from the client through the load balancer is network address translated (NAT) with the load balancer as its source address. The nodes send their return traffic to the load balancer before being passed back to the client. Without this traffic, return traffic goes directly back to the client and connections fail.

In a multiarm configuration, the traffic is routed through the load balancer. The end devices typically have the load balancer as their default gateway.

The most common deployment is a one-arm configuration. The configurations in the figure assumes a one-arm configuration, as this is most commonly deployed. The same principles apply to multiarm deployments, and they both work with F5. For the purpose of this document, the vRealize Automation components are deployed as a one-arm configuration as shown in Figure 1.

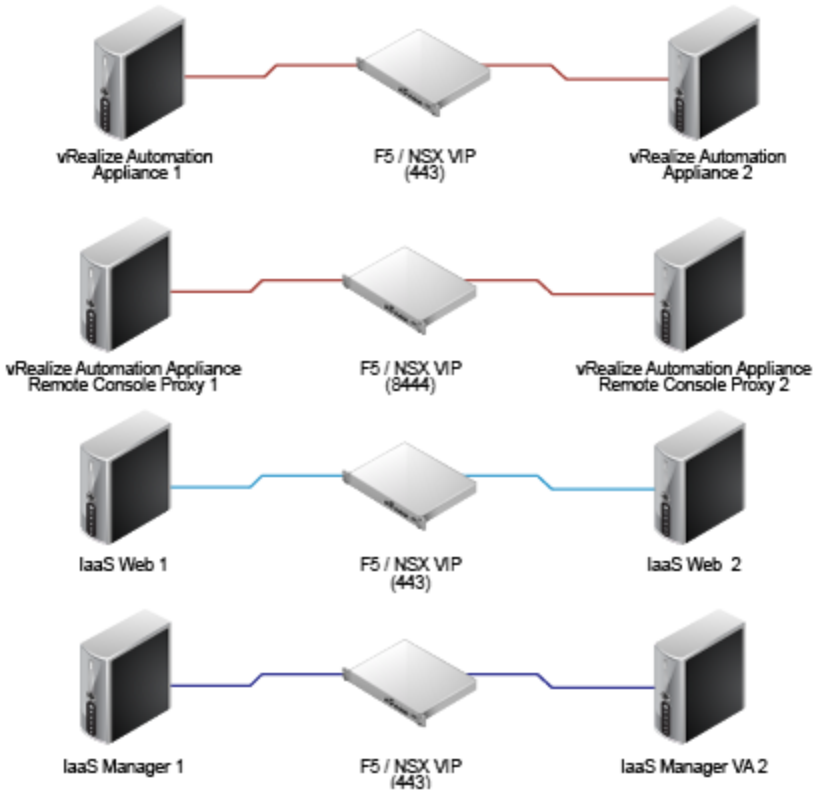


FIGURE 1. ONE-ARM CONFIGURATION

Prerequisites for configuring F5 with vRealize Automation

- **F5 -** Before you start the HA implementation of vRealize Automation using an F5 load balancer, ensure that F5 is installed and licensed and that the DNS server configuration is complete.
- **NSX -** Before you start the HA implementation of vRealize Automation using NSX as a load balancer, ensure that your NSX topology is configured and that your version of NSX is supported. This document covers the load balancing aspect of an NSX configuration, and assumes that NSX is configured and validated to work properly on the target environment and networks.
  To verify that your version is supported, see the *vRealize Automation Support Matrix* for the current release.
- **Certificates -** Create signed or self-signed certificates to contain the vRealize Automation VIP and the hostnames of the vRealize Automation nodes in the SubjectAltNames section. This configuration allows the load balancer to serve traffic without SSL errors. If you need to replace the self-signed certificates with your own CA signed certificates, see VMware Knowledge base article KB 2107816. For more information about certificate troubleshooting and supportability, see the VMware knowledge base article KB 2106583.
- **Identity provider -** With vRealize Automation 7.0, the preferred Identity Provider is VMware Identity Manager which is embedded in the vRealize Automation Appliance
- **Database –** Verify that supported database servers are available for vRealize Infrastructure-as-a-Service (IaaS) nodes. IaaS components require an MS SQL server instance.

For more information on installation and configuration see vRealize Automation 7.0 product documentation.

If required, external Orchestrator cluster can be configured to work with the vRealize Automation system. This can be done after the vRealize Automation system is up and running. However, a vRealize Automation Highly-Available

setup already includes an embedded Orchestrator cluster.

Completing the vRealize Automation Initial Installation

During the initial setup process, the load balancer with all nodes enabled routes half of the traffic to the secondary nodes, which are not yet installed, and the installation fails. To avoid these failures and to complete the initial installation of a vRealize Automation, you must perform the following tasks.

1. Configure the load balancer as described in Configuring F5 Big IP.
2. Turn off the health monitors or change them temporarily to default TCP, and ensure traffic is still forwarding to your primary nodes.
3. Disable all secondary nodes (VA and IaaS) from the load balancer pools.
4. Install and configure all of the system components as detailed in vRealize Automation Installation and Configuration documentation.
5. When all of the components are installed, enable all nodes on the load balancer.
6. Fully configure either the F5 or NSX load balancer with all of the monitors (health-checks) enabled.

After you complete this procedure, update the monitor that you created in Configure Monitors.

Configuring F5 Big IP

This document assumes that the F5 device is already deployed in the environment and is configured with network connectivity to the vRealize Automation components.

- The F5 can be either physical or virtual and can be deployed in one-arm or multiarm topologies
- The Local Traffic module (LTM) must be configured and licensed as either Nominal, Minimum, or Dedicated. You can configure the LTM on the System > Resource Provisioning page

If you are using an F5 version older than 11.x you might need to change your health monitor settings related to the Send string. For more information about how to set up your health monitor send string for the different versions of F5 see HTTP health checks may fail even though the node is responding correctly.

Configure Custom Persistence Profile

You can configure persistence profile for your F5 load balancer by using the following steps.

1. Log in to the F5 and select **Local Traffic** > **Profiles** > **Persistence**.
2. Click **Create**.
3. Enter the name **source_addr_vra** and select **Source Address Affinity** from the drop-down menu.
4. Enable **Custom** mode.
5. Set the **Timeout** to **1800 seconds** (**30 minutes**).
6. Click **Finished**.

Configure Monitors

You can configure required monitors for your F5 load balancer by using the following steps.

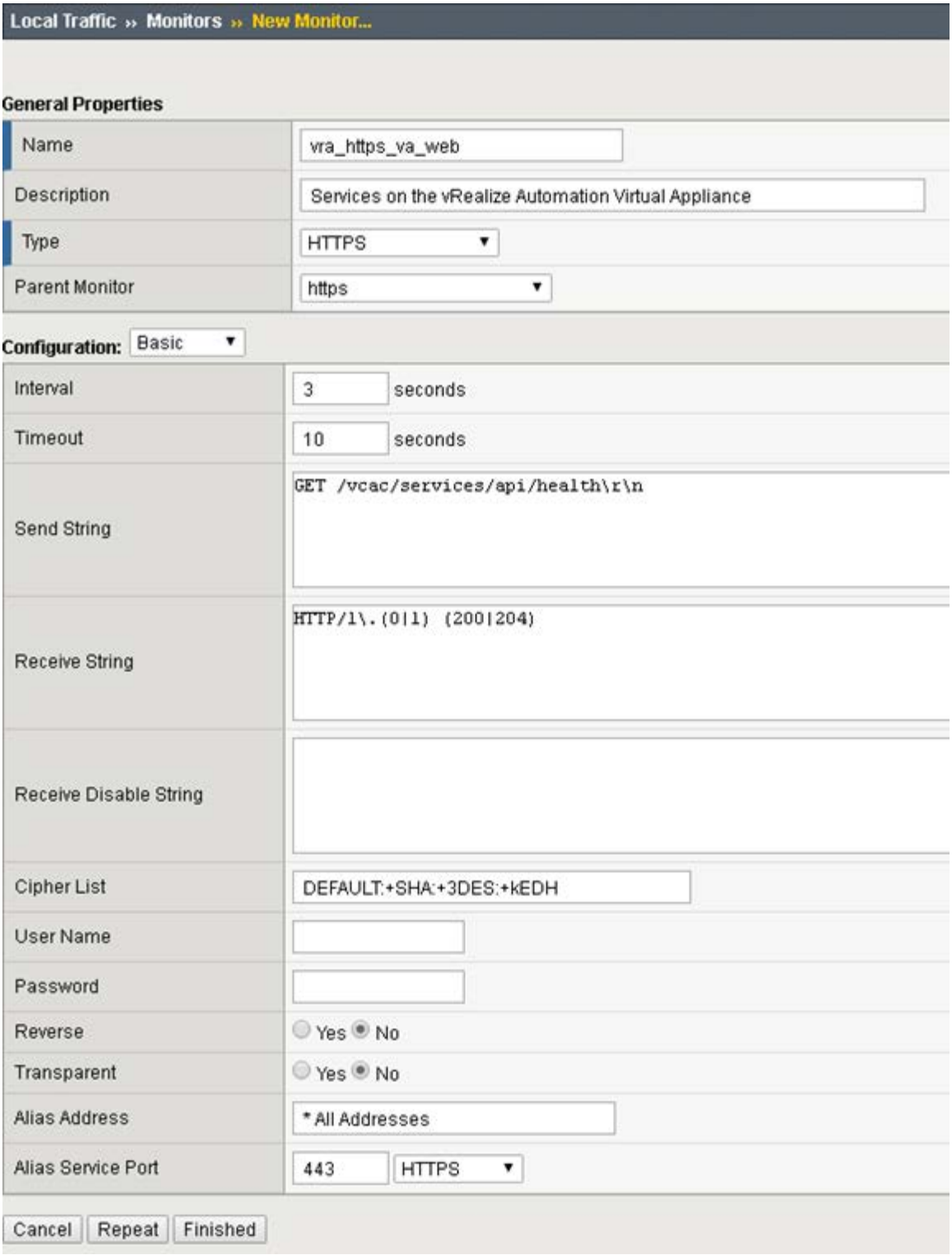
1. Log in to the F5 load balancer and select **Local Traffic** > **Monitors**.
2. Click **Create** and provide the required information. Leave the default when nothing is specified.
3. Repeat steps 1 and 2 for each row of information in Table 1.
4. To check the network map for an overall view of the monitors, select **LTM > Network Map**.

TABLE 1 - CONFIGURE MONITORS

NAME	TYPE	INTERVAL	TIMEOUT	SEND STRING	RECEIVE STRING	ALIAS SERVICE PORT
vra_https_va_web	HTTPS	3	10	GET /vcac/services/api/health\r\n	HTTP/1\.(0|1) (200|204)	443
vra_https_iaas_web	HTTPS	3	10	GET /wapi/api/status/web\r\n	REGISTERED	
vra_https_iaas_mgr	HTTPS	3	10	GET /VMPSProvision\r\n	ProvisionService	

Example

The completed configuration for a VA monitor should look similar to the following screen:



Configure Server Pools

You can configure server pools for your F5 load balancer by using the following steps.

1. Log in to the F5 load balancer and select **Local Traffic** > **Pools**.
2. Click **Create** and provide the required information. Leave the default when nothing is specified.
3. Enter each pool member as a **New Node** and add it to the **New Members**.
4. Repeat steps 1, 2, and 3 for each row of information in Table 2.
5. To check the network map for an overall view of the server pools, select **LTM > Network Map**.

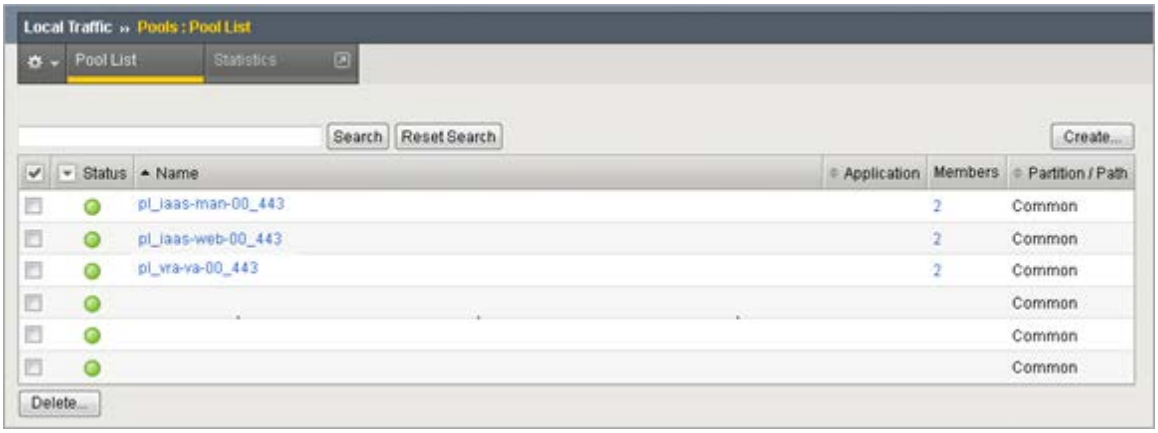
TABLE 2 – CONFIGURE SERVER POOLS

NAME	HEALTH MONITORS	LOAD BALANCING METHOD	NODE NAME	ADDRESS	SERVICE PORT
pl_vra-va-00_443	vra_https_va_web	Round Robin	ra-vra-va-01	10.26.38.44	443
			ra-vra-va-02	10.26.38.45	443
pl_iaas-web-00_443	vra_https_iaas_web	Round Robin	ra-web-01	10.26.38.49	443
			ra-web-02	10.26.38.50	443
pl_iaas-man-00_443	vra_https_iaas_mgr	Round Robin	ra-man-01	10.26.38.46	443
			ra-man-02	10.26.38.59	443
*pl_vra-va-00_8444	vra_https_va_web	Round Robin	ra-vra-va-01	10.26.38.44	8444
			ra-vra-va-02	10.26.38.45	8444

*Port 8444 is optional – it is used for the remote-console connectivity

Example

The completed configuration should look similar to the following screen.



Configure Virtual Servers

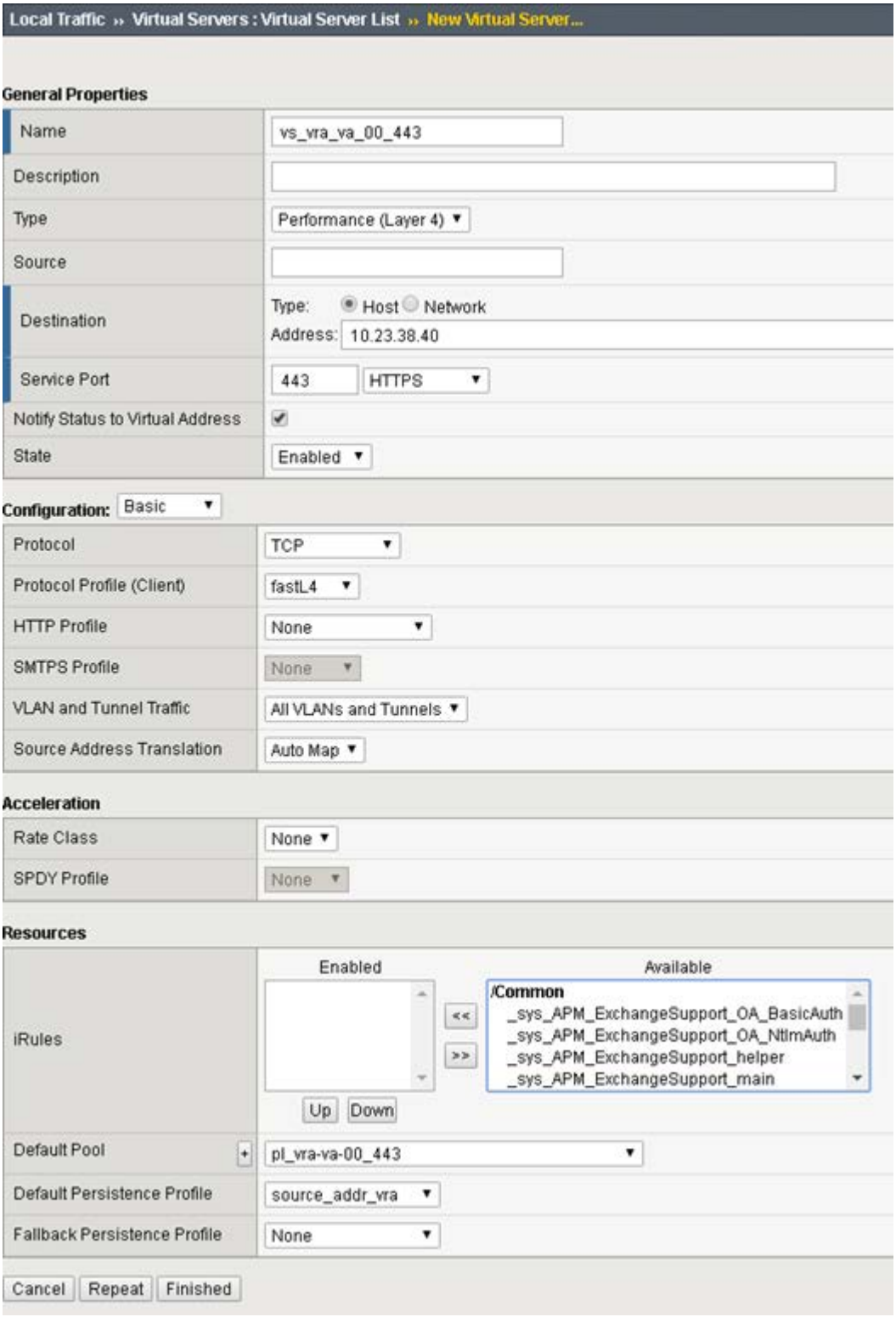
You can configure virtual servers for your F5 load balancer by using the following steps.

1. Log in to the F5 load balancer and select **Local Traffic** > **Virtual Servers**.
2. Click **Create** and provide the required information. Leave the default when nothing is specified.
3. Repeat steps 1 and 2 for each entry in Table 3.
4. To check the network map for an overall view of the virtual servers, select **LTM > Network Map**.

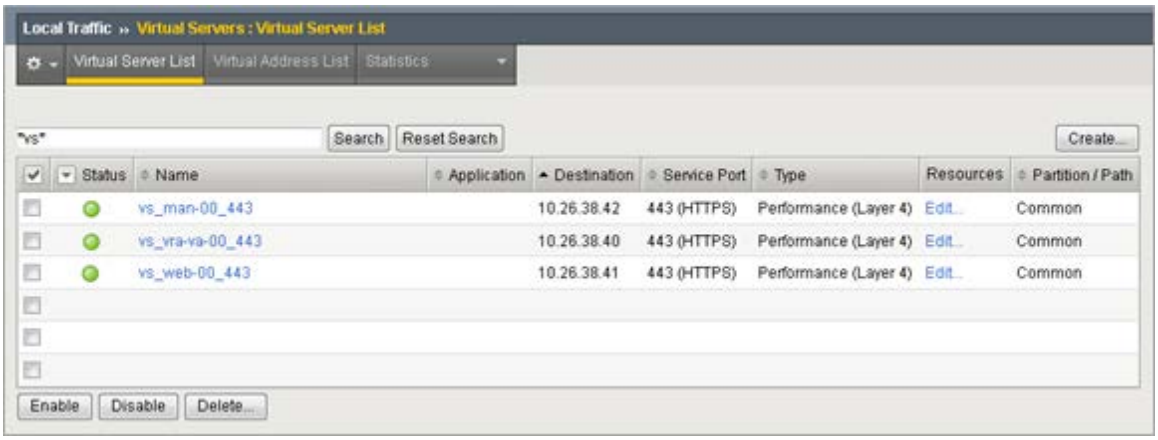
TABLE 3 – CONFIGURE VIRTUAL SERVERS

NAME	TYPE	DESTINATION ADDRESS	SERVICE PORT	SOURCE ADDRESS TRANSLATION	DEFAULT POOL	DEFAULT PERSISTENCE PROFILE
vs_vra-va-00_443	Performance (Layer 4)	10.26.38.40	443	Auto Map	pl_vra-va-00_443	source_addr_vra
vs_web-00_443	Performance (Layer 4)	10.26.38.41	443	Auto Map	pl_iaas-web-00_443	source_addr_vra
vs_man-00_443	Performance (Layer 4)	10.26.38.42	443	Auto Map	pl_iaas-man-00_443	None
vs_vra-va-00_8444	Performance (Layer 4)	10.26.38.40	8444	Auto Map	pl_vra-va-00_8444	source_addr_vra

Example


Local Traffic ›› Virtual Servers : Virtual Server List ›› New Virtual Server...
General Properties
Name
vs_vra_va_00_443
Description
Type
Performance (Layer 4)
Source
Destination
Type: Host Network
Address: 10.23.38.40
Service Port
443
HTTPS
Notify Status to Virtual Address
State
Enabled
Configuration: Basic
Protocol
TCP
Protocol Profile (Client)
fastL4
HTTP Profile
None
SMTPS Profile
None
VLAN and Tunnel Traffic
All VLANs and Tunnels
Source Address Translation
Auto Map
Acceleration
Rate Class
None
SPDY Profile
None
Resources
iRules
Enabled
Available
/Common
_sys_APM_ExchangeSupport_OA_BasicAuth
_sys_APM_ExchangeSupport_OA_NtlmAuth
_sys_APM_ExchangeSupport_helper
_sys_APM_ExchangeSupport_main
Up
Down
Default Pool
pl_vra-va-00_443
Default Persistence Profile
source_addr_vra
Fallback Persistence Profile
None
Cancel
Repeat
Finished

The completed configuration should look similar to the following screen.



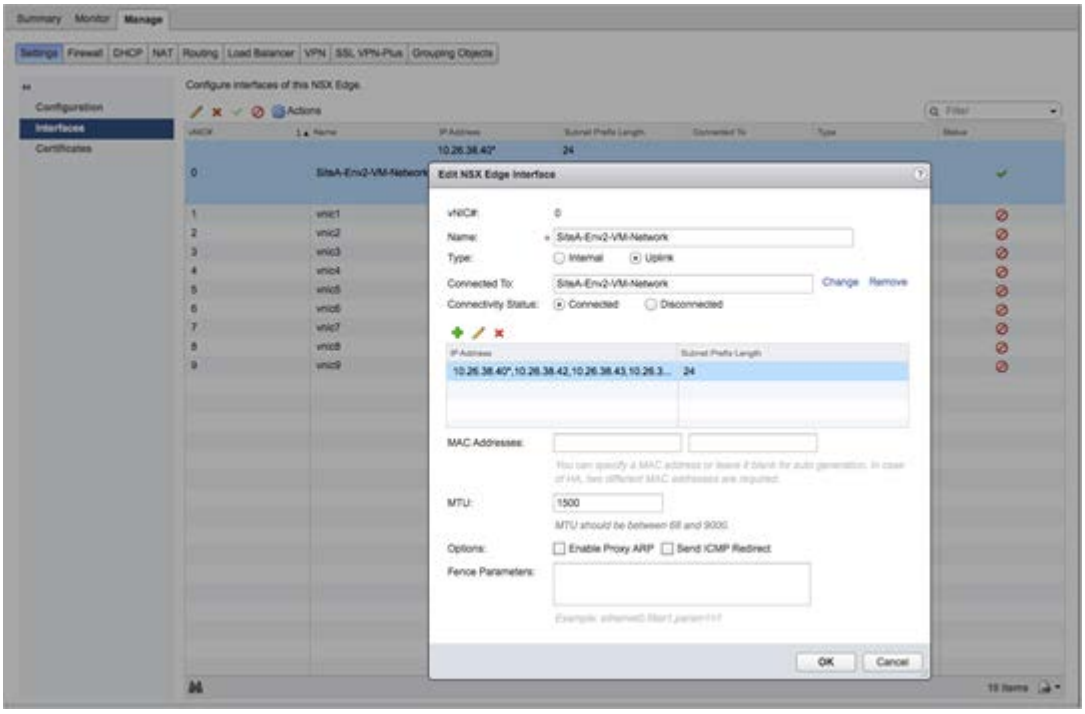
Configuring NSX 6.1.x

You can deploy a new NSX Edge Services Gateway or use an existing one. It must have network connectivity to and from the vRealize Automation components being load balanced.

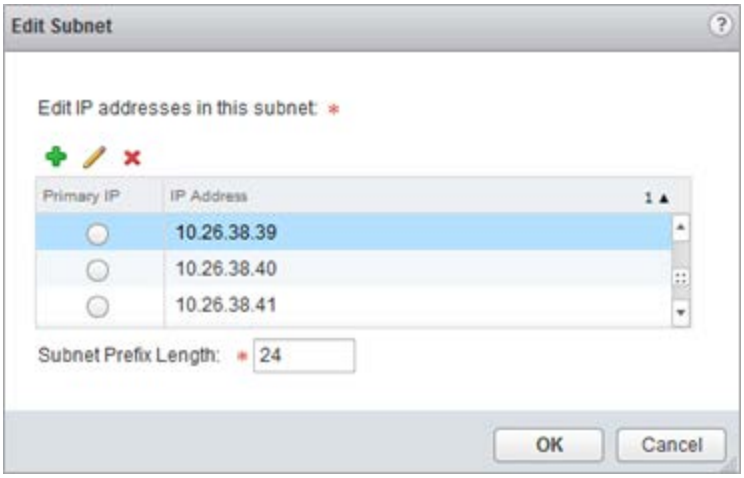
Configure Global Settings

You can configure the global settings by using the following steps.

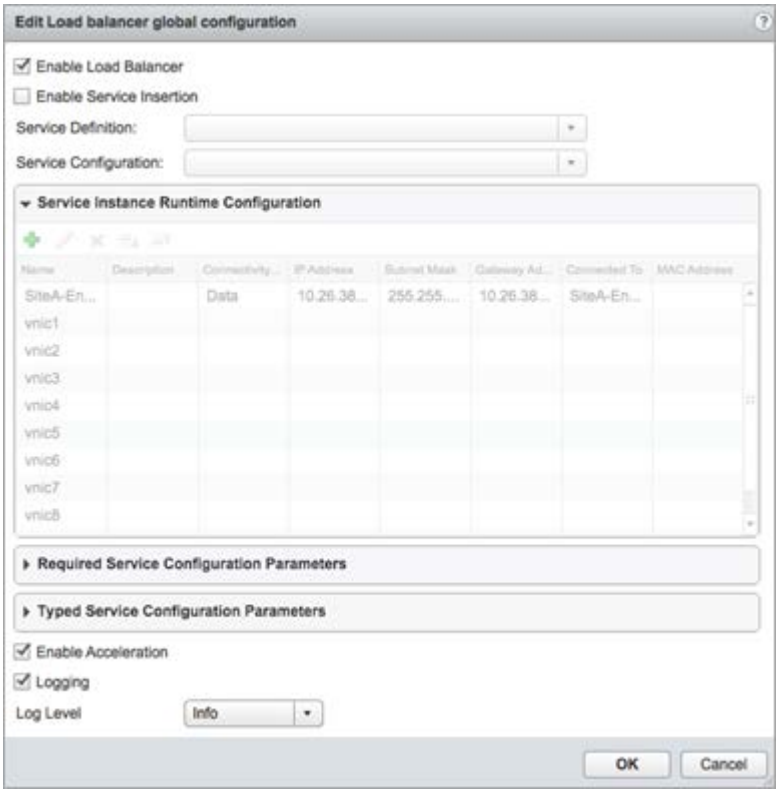
1. Log in to the NSX, select the **Manage** tab, click **Settings**, and select **Interfaces**.
2. Double-click to select your Edge device from the list.
3. Click **vNIC#** for the external interface that hosts the VIP IP addresses and click the **Edit** icon.
4. Select the appropriate network range for the NSX Edge and click the **Edit** icon.



5. Add the IP addresses assigned to the VIPs, and click **OK**.
6. Click **OK** to exit the interface configuration subpage.



7. Select the **Load Balancer** tab and click the **Edit** icon.
8. Select **Enable Load Balancer**, **Enable Acceleration**, and **Logging**, if required, and click **OK**.



Add Application Profiles

You can add application profiles for different components of vRealize Automation.

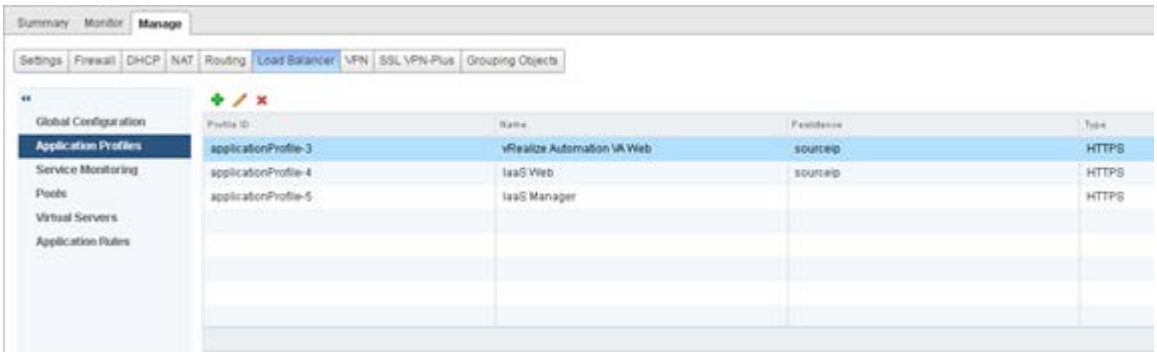
1. Click **Application Profiles** on the window pane on the left.
2. Click the **Add** icon to create the Application Profiles required for vRealize Automation by using information in Table 4. Leave the default when nothing is specified.

TABLE 4 – ADD APPLICATION PROFILES

NAME	TYPE	ENABLE SSL PASS-THROUGH	TIMEOUT	PERSISTENCE
IaaS Manager	HTTPS	Checked	-	None
IaaS Web	HTTPS	Checked	1800 seconds	Source IP
vRealize Automation VA Web	HTTPS	Checked	1800 seconds	Source IP

Example

The completed configuration should look similar to the following screen.



Add Service Monitoring

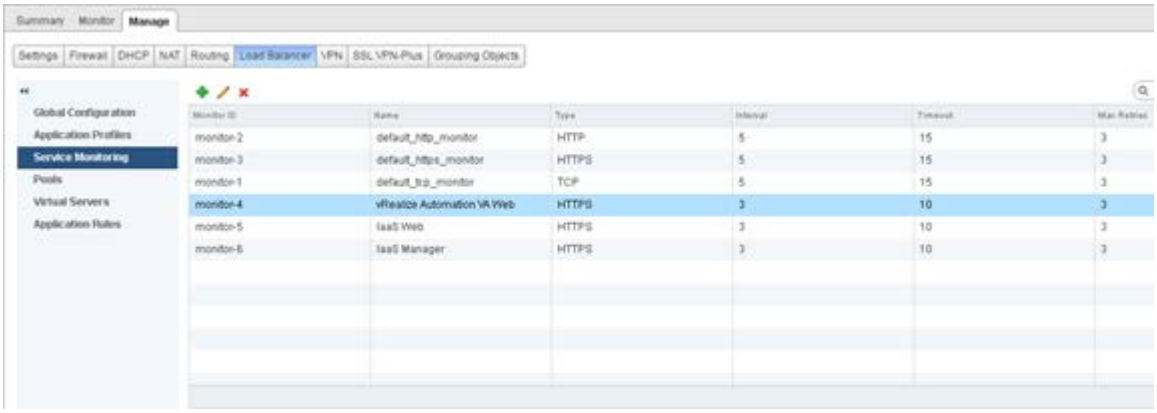
You can add service monitoring for different components of vRealize Automation.

1. Click **Service Monitoring** in the left pane.
2. Click the **Add** icon to create the Service Monitors required for vRealize Automation using information in Table 5. Leave the default when nothing is specified.

TABLE 5 – ADD SERVICE MONITORING

NAME	INTERVAL	TIME OUT	RETRIES	TYPE	METHOD	URL	RECEIVE	EXPECTED
vRealize Automation VA Web	3	10	3	HTTPS	GET	/vcac/services/api/health		200, 204 (for 7.0) 204 (for 7.0.1 and higher)
IaaS Web	3	10	3	HTTPS	GET	/wapi/api/status/web	REGISTERED	
IaaS Manager	3	10	3	HTTPS	GET	/VMPSProvision	ProvisionService	

The completed configuration should look similar to the following screen.



Add Pools

You can create pools by using the following steps.

1. Click **Pools** in the left pane.
2. Click the **Add** icon to create the Pools required for vRealize Automation using information in Table 6. Leave the default when nothing is specified.

You can either use the IP address of the pool members, or select them as a Virtual Center Container.

TABLE 6 - ADD POOLS

POOL NAME	ALGORITHM	MONITORS	MEMBER NAME	EXAMPLE IP ADDRESS / VCENTER CONTAINER	PORT	MONITOR PORT
pool_vra-va-web_443	Round Robin	vRA VA Web	vRA VA1	10.26.38.44	443	
			vRA VA2	10.26.38.45	443	
pool_iaas-web_443	Round Robin	IaaS Web	IaaS Web1	10.26.38.49	443	
			IaaS Web2	10.26.38.50	443	
pool_iaas-manager_443	Round Robin	IaaS Manager	IaaS Man1	10.26.38.49	443	
			IaaS Man2	10.26.38.50	443	
*pool_vra-rconsole_8444	Round Robin	vRA VA Web	vRA VA1	10.26.38.44	8444	443
			vRA VA2	10.26.38.45	8444	443

*Only needed if remote-console access is used

Add Virtual Servers

You can add virtual servers by using the following steps.

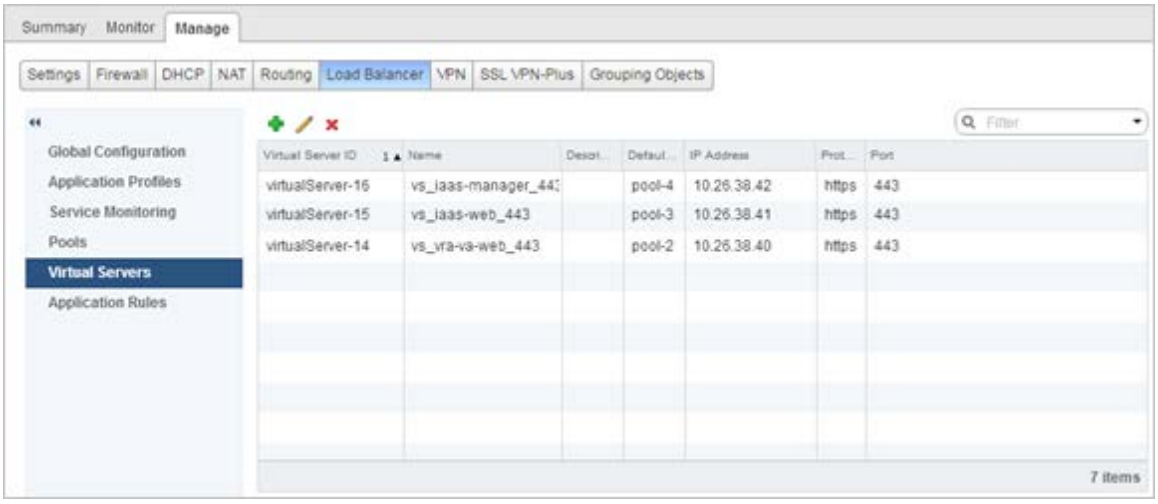
1. Click **Virtual Servers** on the left pane.
2. Click the **Add** icon to create the Virtual Servers required for vRealize Automation using the information in Table 7. Leave the default when nothing is specified.

TABLE 7 - ADD VIRTUAL SERVERS

NAME	IP ADDRESS	PROTOCOL	PORT	DEFAULT POOL	APPLICATION PROFILE	APPLICATION RULE
vs_vra-va-web_443	10.26.38.40	HTTPS	443	pool_vra-va-web_443	vRA VA	
vs_iaas-web_443	10.26.38.41	HTTPS	443	pool_iaas-web_443	IaaS Web	
vs_iaas-manager_443	10.26.38.42	HTTPS	443	pool_iaas-manager_443	IaaS Manager	
*vs_vra-va-rconsole_8444	10.26.38.40	HTTPS	8444	pool_vra-rconsole_8444	vRA VA	

*Only needed if remote-console access is used

The completed configuration should look similar to the following screen.



VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com