Avi Vantage Configuration Guide

Virtual Services 9
Virtual Service Page 10
Virtual Services Details Pages 12
Virtual Service Quick Info Popup 13
Virtual Services Analytics 14
About Virtual Service Analytics 14
End-to-End Timing 15
Metrics Tiles 17
Chart Pane 18
Overlays Pane 18
Anomalies Overlay 19
Alerts Overlay 21
Configuration Events Overlay 22
System Events Overlay 23
Virtual Service Logs 24
Enabling Logs 24
Significant Logs 24
Full Client Logs 24
Rotation Out of Unrequested Logs 25
Logging During Heavy Load 25
Log Navigation 25
Log Analytics 28
Log Analytics Tiles 29
Expanded Logs 31
Searching Logs 34
Significant Log Events 35
Virtual Service Health Monitoring 38
Displaying Virtual Service Health 38
Summary Health Scores 38
Health Details 39
Health Score Codes 40
Health Score Composition 40
Virtual Service Client Information 42
Resource Timing Metrics 45
Virtual Service Security 47
View Security Insights 47
SSL Distribution Insights 48
SSL Insights 49
DDoS Insights 50
Virtual Service Events Log 51
Virtual Service Events 51
Virtual Service Alerts 53
Alert Actions 54
Alert Table 54
Create a Virtual Service 55
Virtual Service Creation: Basic Mode 55
Virtual Service Creation: Advanced Mode 58

Advanced Virtual Service Creation: Settings Tab 58
Advanced Virtual Service Creation: Rules Tab 62
Policy Types 63
Policy Processing Order 63
Adding a Policy 63
DataScript 64
Advanced Virtual Service Creation: Analytics Tab 65
Analytics 65
Client Insights 66
Non-significant Logs 66
Advanced Virtual Service Creation: Advanced Tab 67
Performance Limits 67
Quality of Service 69
Placement Settings 69
HTTP Basic Authentication 69
Other Settings 70
Virtual Service Scaling 71
Operational Notes 71
How Scaling Operates in VMware / OpenStack with Nuage Deployments 71
How Scaling Operates in OpenStack with Neutron Deployments 72
How High Availability Affects Scaling 72
Scaling Process 72
Scaling Out 73
Scaling In 74
Migrating 74
Virtual Service Policies 74
Prioritizing Policies 75
Match / Action 75
Create a Policy 76
Network Security 78
HTTP Security 78
Policy Tokens 80
Pools 82
Pools Page 83
Pool Details Page 84
Pool Analytics Page 84
Pool End-to-End Timing 85
Pool Metrics 86
Pool Chart Pane 87
Pool Overlays Pane 88
Pool Anomalies Overlay 89
Pool Alerts Overlay 90
Pool Config Events Overlay 91
Pool System Events Overlay 92
Pool Logs Page 92
Pool Health Page 93
Pool Servers Page 94
Server Page 94
Create Pool 99
Create Pool: 1 Settings 99
Create Pool: 2 Servers 103
Add Servers 103
Servers 104
Create Pool: 3 Advanced 106
Create Pool: 4 Review 108
Servers 108
Server Page 109
Alerts 111
All Alerts 111

Alerts for Specific Virtual Service 111
Alert Configuration 112
Alert Config Page 113
Create Alert Configuration 113
Alert Actions 115
Alert Action Page 116
Create/Edit an Alert Action 117
Events 118
All Events 118
Config Audit Trail 119
Services 121
Syslog Notifications 121
Syslog Notifications Settings 121
Create Syslog Notification 122
Email 122
SNMP Trap 123
SNMP Trap Settings 123
Traffic Capture 124
Capture Configuration 124
Completed Captures 125
Templates 125
Profiles 126
Application Profile 126
Dependency on TCP/UDP Profile 126
Application Profile Tab 127
Create/Edit an Application Profile 127
HTTP Profile 128
HTTP General Tab 128
Multiplex plus Persistence 129
HTTP Security 130
Security Information 130
Client SSL Certificate Validation 132
HTTP Acceleration 132
HTTP Compression 133
Custom Compression 134
HTTP Front-end Optimization 135
HTTP Caching 135
HTTP DDoS 137
HTTP Limits 138
Rate Limits 139
DNS Profile 141
L4 Profile 141
Syslog Profile 141
TCP/UDP Profile 142
TCP/UDP Profiles Settings 142
Create TCP/UDP Profile 143
TCP Proxy 144
TCP Fast Path 146
UDP Fast Path 147
Persistence Profile 148
Persistence Profile Settings 149
Create Persistence Profile 149
Health Monitor Profile 151
Health Monitor Settings 152
Active Health Monitor 153
DNS Monitor 155

External Monitor 155
HTTP Monitor 156
HTTPS Monitor 157
Ping Monitor 158
TCP Monitor 158
UDP Monitor 158
Passive Health Monitor 159
Analytics Profile 159
Analytics Settings 160
Create an Analytics Profile 160
Groups 164
IP Group 164
Create an IP Group 165
String Group 166
Create a String Group 167
MicroService Group 167
Security 168
SSL Certificates 168
SSL/TLS Certificates Page 168
Create Certificate 169
Self-Signed Certificates 170
CSR Certificates 172
Import Certificates 173
Certificate Authority 175
SSL Profile 175
SSL Profile Settings 176
Create an SSL Profile 176
PKI Profile 177
PKI Profile Settings 178
Create a PKI Profile 178
Authentication Profile 179
Auth Profile Settings 180
Create an Authentication Profile 180
Scripts 182
DataScript 183
ControlScript 183
Infrastructure 184
Dashboard 184
Clouds 184
Cloud Creation 186
Cloud Management 186
Service Engines 187
Service Engines Page 187
Service Engines Details Page 188
Service Engine Quick Info 190
Service Engine Analytics 191
Service Engine Analytics: Metrics 191
Service Engine Analytics: Chart Pane 192
Service Engine Analytics: Overlays Pane 193
SE Analytics: Anomalies Overlay 194
SE Analytics: Alerts Overlay 195
SE Analytics: Config Events Overlay 196
SE Analytics: System Events Overlay 197
Service Engine Health 198
Service Engine Events 198
Service Engine Alerts 200
Service Engine Create: Write Access Mode Deployments 202
Service Engine Create: Read/No Access Mode Deployments 202

Edit a Service Engine 203
Delete a Service Engine 205
Service Engine Group 205
Service Engine Groups Page 206
Create a Service Engine Group 207
High Availability Tab 207
Virtual Service High Availability Settings 207
Service Engine Capacity and Limit Settings 210
Advanced Service Engine Group Settings 211
Service Engine Group Network Settings 213
Edit Service Engine Group Network Settings 214
Static Route 216
Static Routes Tab 216
Create/Edit Static Route 217
Infrasture 217
Users 217
Users Settings 218
Create a User 219
User Preferences 220
Roles 221
Roles Settings 222
Create a Role 223
Tenants 224
Tenants Settings 225
Create a Tenant 225
Settings 226
Access Settings Tab 226
DNS / NTP Settings 226
Update DNS/NTP Settings 227
License Tab 228
Email/SMTP Tab 230
Controller 230
Controller Table 230
Controller Analytics 231
Controller Events 232
Controller Alerts 233
Converting a Single-node Cluster to a Three-node Cluster 233
High Availability Operation for Controllers 235
Quorum 236
Failover 236

Avi Vantage Configuration Guide

Architecture

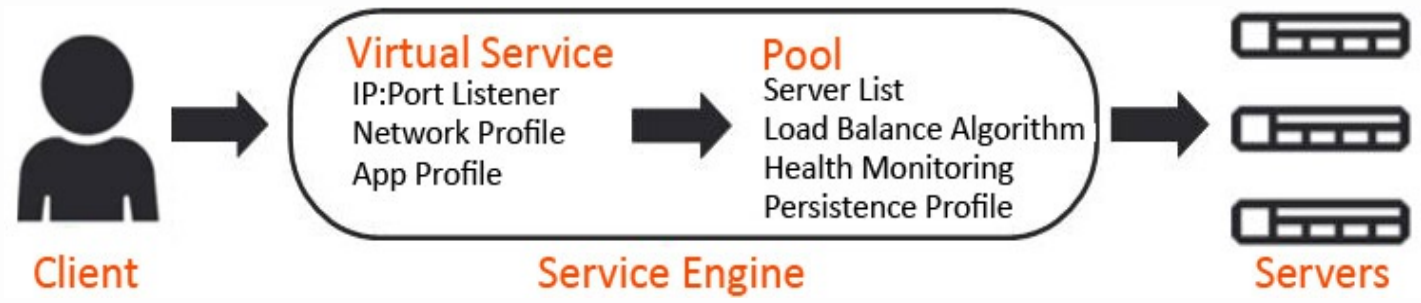
Avi Vantage is built using a modern software defined network (SDN) architecture that centralizes the control plane while distributing the data plane. This allows network services to scale near infinitely without the added complexity of managing hundreds of disparate appliances.

Avi Vantage consists of the following key components:

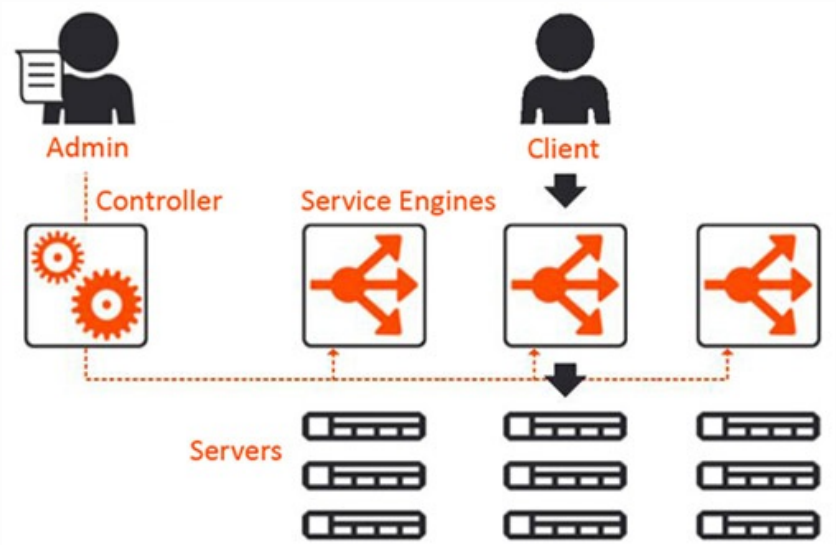
Controller: The Controller is a virtual machine that functions as the manager and single point of access for configuration, visibility, and logs for each Avi Service Engine (SE). Avi Vantage supports three Controllers in a redundant cluster to provide scale and fault tolerance. The Controller is the single point of management for Avi Vantage, via the web interface, REST API, or CLI. By default, the web interface is accessible via HTTPS. (This is configurable.) Internet Explorer (IE) version 11 or later, or any reasonably recent version of Firefox, Safari, Opera or Chrome is supported.

Service Engine: Avi SEs handle all data plane operations within Vantage by receiving and executing instructions from the Controller. The SEs perform load balancing and all client- and server-facing network interactions. High availability is supported.

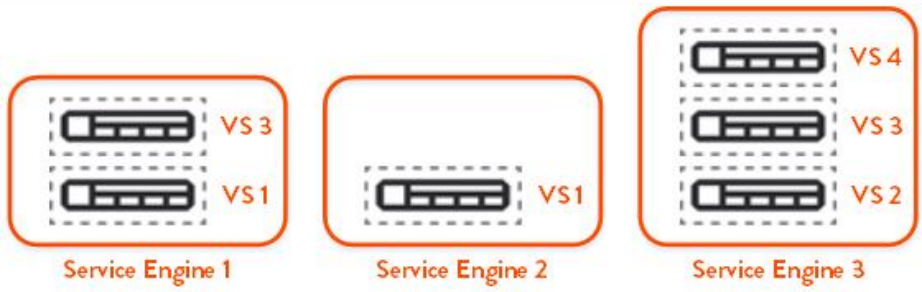
In a typical load balancing scenario, a client will communicate with a virtual service, which is an IP address and port hosted in Vantage by an SE. The virtual service internally passes the connection through a number of profiles. For HTTP traffic, the SE may terminate and proxy the client TCP connection, terminate SSL, and proxy the HTTP request. Once the request has been validated, it will be forwarded internally to a pool, which will choose an available server. A new TCP connection then originates from the SE, using an IP address of the SE on the internal network as the request's source IP address. Return traffic takes the same path back. The client communicates exclusively with the virtual service IP address, not the real server IP.



The Controller managing the Vantage system is typically a redundant three-node cluster. The entire Vantage system is managed through a centralized point (and IP address) regardless of the number of new applications being load balanced and the number of SEs required to handle the load. The Vantage web interface provides visibility into all applications configured. Controllers can automatically create and configure new SEs as new applications are configured via virtual services (in write access mode deployments).



Controllers continually exchange information securely with the SEs and with one other. The SEs perform data plane operations such as load balancing. The SEs also report server health, client connection statistics, and client-request logs to the Controllers, which share the processing of the logs and analytics information. The Controllers also send commands down to the SEs, such as configuration changes. Controllers and SEs communicate over their management IP addresses. (Click here for a list of the protocol ports Vantage uses for management.)



Virtual services may be scaled across one or more SEs. When a virtual service is scaled across multiple SEs, each of those SEs share the load. This sharing may not be equal, because the actual workload distribution depends on the available CPU and other resources that may be required of the SE. SEs typically process traffic for more than one virtual service at a time.

One SE will be the primary for a given virtual service and will advertise that virtual service's IP address from the SE's own MAC address. The primary SE may either process and load balance a client connection itself, or it may forward the connection via layer 2 to the MAC address of another SE that has available capacity.

Each SE will load balance and forward traffic to servers using its own IP address within the server network as the source IP address of the client connection. This ensures that even though multiple SEs may be sending traffic to the same pool of servers, return traffic takes the same path from the servers back through the same SE. When deployed in a VMware environment and the SEs

are scaled out, the secondary SEs will respond directly back to clients without passing the return traffic back through the primary SE. In OpenStack, the secondary SEs will return client responses back through the primary SE.

System Requirements: Ecosystem

Hypervisor Support:

- VMware vSphere 5.1, 5.5, 6.0
- KVM - RHEL/CentOS 6.4, Ubuntu 12.04, 14.04
- Amazon Web Services (AWS)
- Docker containers

VMware Orchestrator Support:

- VMware vCenter 5.1, 5.5, 6.0
- OpenStack Version Support: Havana, Icehouse, Juno, Kilo

OpenStack Vendor Support:

- Trunk
- RedHat
- Mirantis
- Canonical/Ubuntu

Cisco APIC Support:

- Version 1.03(f) or later

CloudStack Version Support:

- CloudStack 4.5.1

Avi Vantage may be deployed in various environments with write (recommended), read, or no access integration with the virtualization orchestrator. The primary difference among these modes is the level of automation performed by Vantage and the cloud orchestrator compared to the level of manual configuration required of administrators. There are no differences in hardware or system requirements among these modes. Servers being load balanced by Avi Vantage may exist within the same virtualization environment or be bare-metal, non-virtualized servers.

Avi supports the ability to manage multiple cloud environments from a single Controller cluster.

System Requirements: Hardware

Avi Vantage runs on standard x86-based servers, with no requirement for special-purpose hardware. In general, adding hardware capacity will greatly expand overall system capacity, for both SEs and Controllers. Please consult an Avi sales engineer or Avi technical support for recommendations tailored to meet the specific needs of your applications and environment.

The minimum hardware requirements are:

- **Controller:** 4 vCPU cores, 12 GB RAM, and 64 GB of storage.
- **Service Engine:** 2 vCPU cores, 2 GB RAM, and 10 GB of storage.

A typical deployment will have three Controllers in a redundant Controller cluster. The number of SEs required will depend on the number of applications being served by Avi Vantage and the configured level of redundancy.

Notes:

- Reservation for CPU and memory is strongly preferred, but not required.
- Modifying resource settings on VMs, such as CPU cores or RAM, requires powering down the VM, making the changes, and then powering the VM back on.

Virtual Services

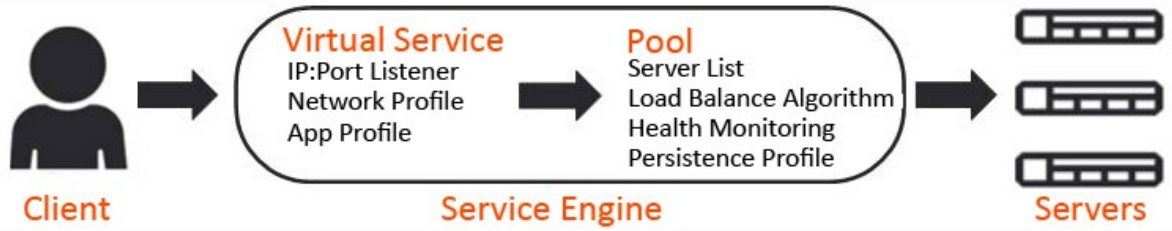
Virtual services are the core of the Vantage Vantage load-balancing and proxy functionality. A virtual service advertises an IP address and ports to the external world and listens for client traffic. When a virtual service receives traffic, it may be configured to:

- Proxy the client’s network connection.
- Perform security, acceleration, load balancing and other tasks.
- Forward the client’s request data to the destination pool for load balancing.

A virtual service can be thought of as an IP address that Vantage is listening to, ready to receive requests. In a normal TCP/HTTP configuration, when a client connects to the virtual service address, Vantage will process the client connection or request against a list of settings, policies and profiles, then send valid client traffic to a back-end server that is listed as a member of the virtual service’s pool.

Typically, the connection between the client and Vantage is terminated or proxied at the SE, which opens a new TCP connection between itself and the server. The server will respond back directly to the Vantage IP address, not to the original client address. Vantage forwards the

response to the client via the TCP connection between itself and the client.



A typical virtual service consists of a single IP address and service port that uses a single network protocol. Vantage allows a virtual service to listen to multiple service ports or network protocols.

For instance, a virtual service could be created for both service port 80 (HTTP) and 443 SSL (HTTPS). In this example, clients can connect to the site with a non-secure connection and later be redirected to the encrypted version of the site. This allows administrators to manage a single virtual service instead of two. Similarly, protocols such as DNS, RADIUS and Syslog can be accessed via both UDP and TCP protocols.

It is possible to create two unique virtual services, where one is listening on port 80 and the other is on port 443; however, they will have separate statistics, logs, and reporting. They will still be owned by the same Service Engines (SEs) because they share the same underlying virtual service IP address.

To send traffic to destination servers, the virtual service internally passes the traffic to the pool corresponding to that virtual service. A virtual service normally uses a single pool, though an advanced configuration using policies or DataScripts can perform content switching across multiple pools. A script also can be used in lieu of a pool, such as a virtual service that only performs an HTTP redirect.

A pool can only be assigned to a single virtual service. If the virtual service is deleted or pointed at a different pool, the pool will become unassigned and available to be used by a different virtual service.

When creating a virtual service, that virtual service listens to the client-facing network, which is most likely the upstream network where the default gateway exists. The pool connects to the server network.

Normally, the combined virtual service and pool are required before Vantage can place either object on an SE. When making an SE placement decision, Vantage must choose the SE that has the best reachability or network access to both client and server networks. Alternatively, both the clients and servers may be on the same IP network.

Virtual Service Page

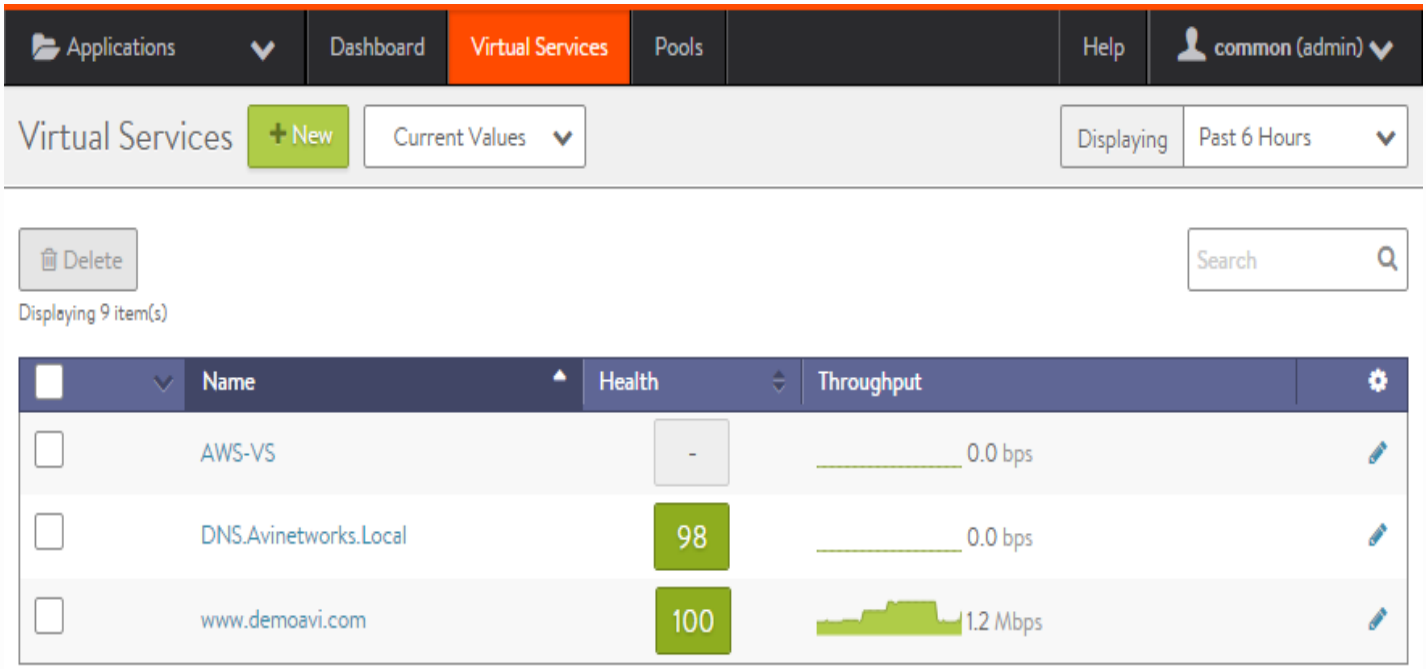
Select Applications > Virtual Services to open the virtual services page. This page displays a list

of the configured virtual services. It can be used to quickly check the status and view high level information for each.

This page includes the following functions:

- **Search:** Search through the names of the objects.
- **Create:** Opens the *Create Virtual Service* popup.
- **Edit:** Opens the *Edit Virtual Service* popup.
- **Delete:** Immediately removes a virtual service from Vantage. This will terminate all existing connections, delete the configuration of the virtual service, and place the pools used by that virtual service into an unused state. When deleting, a secondary prompt will ask to remove the pool at the same time or leave it intact. The SEs hosting the virtual service may be eligible for deletion if they are no longer in use. Note that an object cannot be un-deleted.

The table on this page contains the following information for each virtual service. The columns are customizable, so the exact view may be different.



- **Name:** Lists the name of each virtual service. Clicking the name of a virtual service opens the Analytics tab of the Virtual Service Details page.
- **Health:** Provides both a number from 1-100 and a color-coded status to provide quick information about the health of each virtual service. If the virtual service is down, an exclamation point will appear instead of a number. A dash appears if the virtual service is disabled, not deployed, or in an error state.
  - Hovering the cursor over this score opens the Health Score popup for the virtual service.

    - The View Insights link at the bottom of the popup opens the Insights tab of the Virtual Service Details page.
    - Clicking within the Health Score popup opens the Analytics tab of the Virtual Service Details page.
- **Address:** Displays the IP address advertised by the virtual service.
- **Services:** Lists the service ports configured for the virtual service. Ports that are configured for terminating SSL/TLS connections are denoted in parenthesis. A virtual service may have multiple ports configured. For example:
    - 80 (HTTP)
    - 443 (SSL)
- **Pools:** Lists the pools assigned to each virtual service. Clicking a pool name opens the Analytics tab of the Pool Details Page.
- **Service Engines:** Lists the Service Engines that the virtual server is assigned to. Clicking a Service Engine name opens the Analytics tab of the Service Engine Details page.
- **# Service Engines**: Shows the number of SEs assigned to the virtual service as a time series. Useful to see if a virtual service scales up or down the number of SEs.
- **Throughput:** Thumbnail chart of the throughput for each virtual service for the time frame selected.
    - Hovering the cursor over this graph shows the throughput for the highlighted time.
    - Clicking a graph opens the Analytics tab of the Virtual Service Details page for the virtual service.

Virtual Services Details Pages

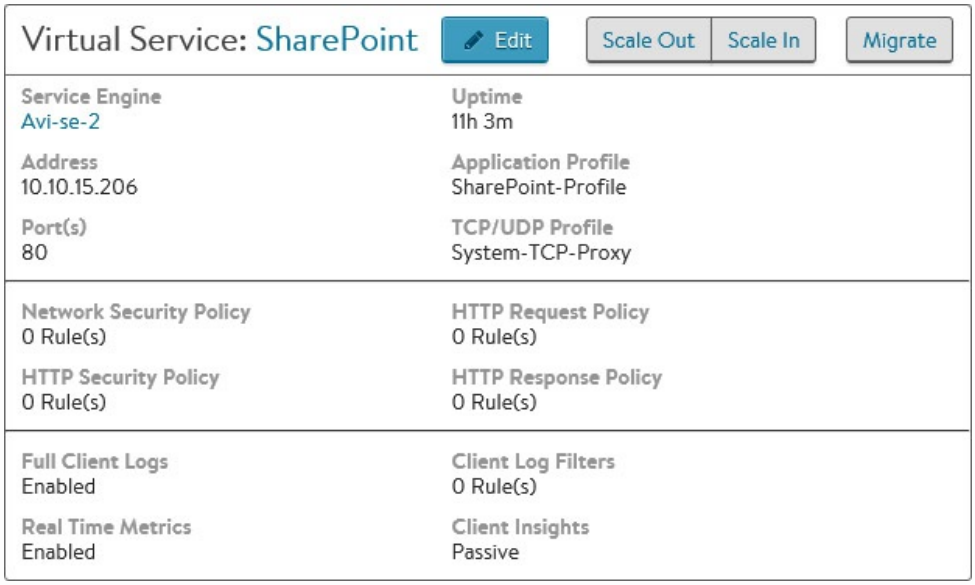
The Virtual Service Details pages shows extensive information about a virtual service. Access these pages by clicking the name of a virtual service within the *Applications* > *Dashboard* or from the *Applications* > *Virtual Service* page.

The details pages are a loose collection of a number of sub-pages under the umbrella of the virtual service.

- Virtual Service Analytics Page
- Virtual Service Logs Page
- Virtual Service Health Page
- Virtual Service Clients Page
- Virtual Service Security Page
- Virtual Service Events Page
- Virtual Service Alerts Page

Virtual Service Quick Info Popup

All of the virtual service details pages include the Virtual Service Quick Info popup, which may be accessed by hovering over or clicking the name of the virtual service in the top left corner of the page.



The Virtual Service Quick Info popup provides buttons for the following functions:

- **Edit:** Opens the Edit Virtual Service popup.
- **Scale Out:** Scales out, which distributes connections for the virtual service to one additional SE per click, up to the maximum number of SEs defined in the SE group properties.
- **Scale In:** Scales in the virtual service by one SE, down to a minimum of one SE.
- **Migrate:** Moves the virtual service from the SE it is currently on to a different SE.

This popup also displays the following information (if applicable) for the virtual service:

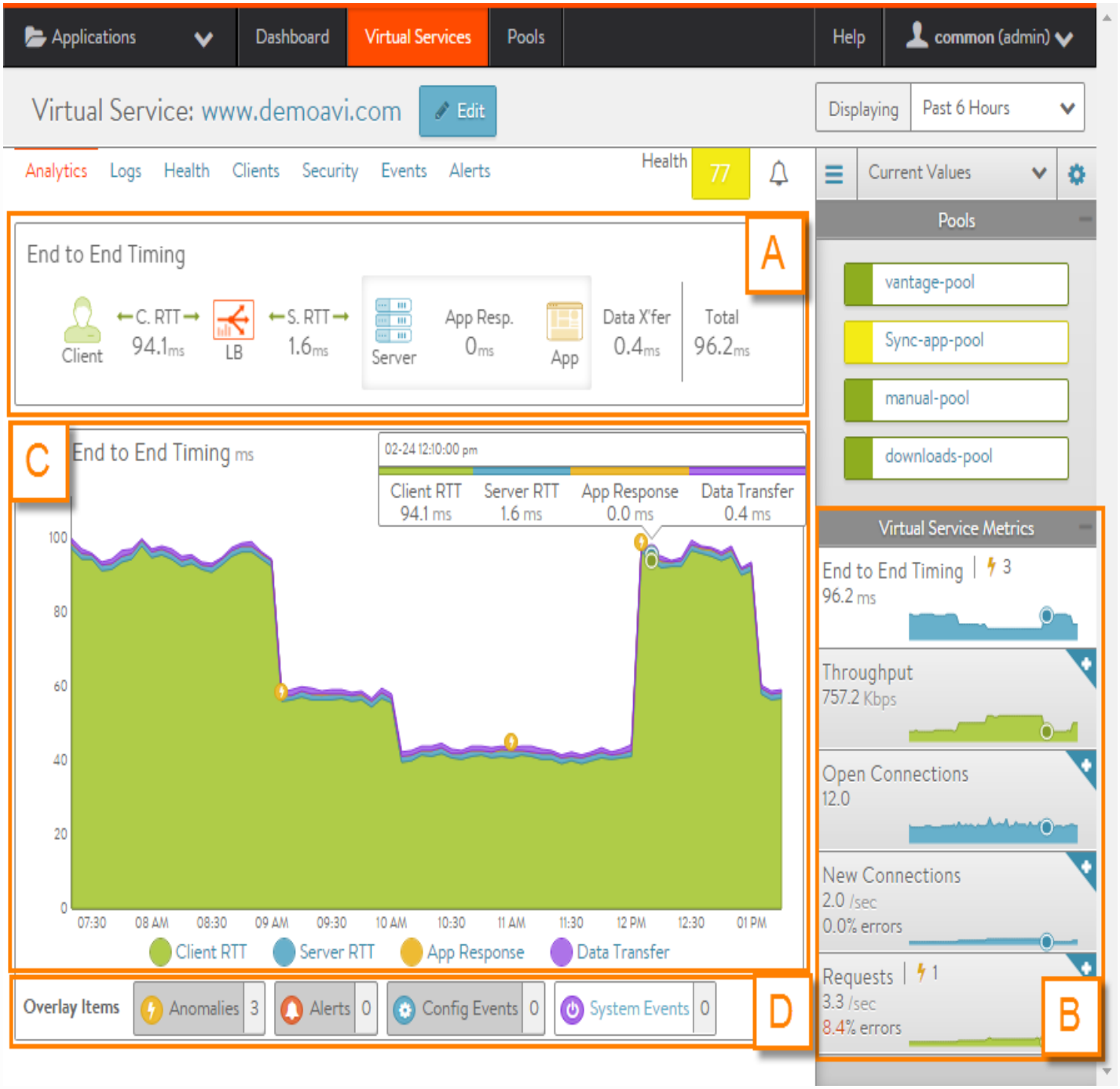
- **Service Engine:** Names of the SEs this virtual service is deployed on. Clicking on an SE name opens the Service Engine Details page for that SE.
- **Uptime / Downtime:** The amount of time the virtual service has been in the current up or down state.
- **Address:** IP address of the virtual service.
- **Application Profile:** The application profile applied to the virtual service.
- **Service Port:** Service ports on which the virtual service is listening for client traffic.
- **TCP/UDP Profile:** The profile applied to the virtual service.

- **Rules:** The number of policies or DataScripts attached to the virtual service.
- **Non-Significant Logs:** When disabled, the virtual service defaults to logging significant events or errors. When enabled, all connections or requests are logged. (The Analytics page has additional logging options.)
- **Real Time Metrics:** When this option is disabled, metrics are collected every five minutes, regardless of whether the Display Time is set to the Real Time. When the option is enabled, metrics are collected every 15 seconds.
- **Client Log Filters:** Number of custom log filters applied to the virtual service. Log filters can selectively generate non-significant or more verbose logs.
- **Client Insights:** Type of client insights gathered by the virtual service: Active, Passive, or None.

Virtual Services Analytics

About Virtual Service Analytics

The VS analytics tab presents information about the virtual service performance metrics. All charts and metrics reflect the display time selected.



This article is broken down into the following sections, outlined in the graphic above.

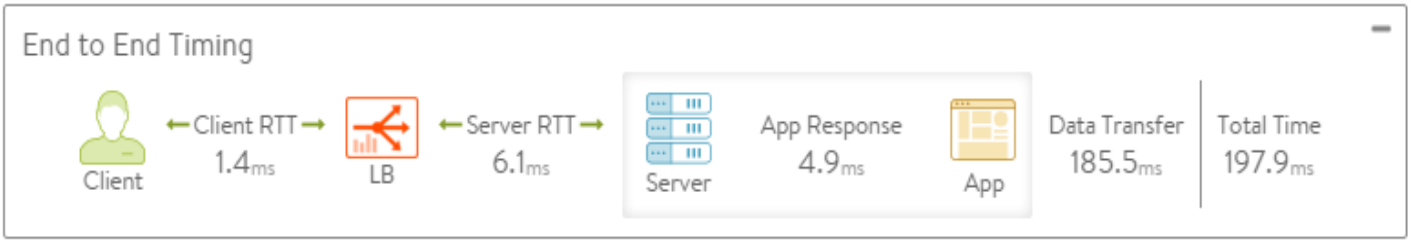
- End-to-End Timing (A)
- Metrics tiles (B)
- Chart pane (C)
- Overlay pane (D)

End-to-End Timing

The End to End Timing pane provides a high-level overview of the quality of the end-user

experience and where any slowdowns may be occurring. The chart breaks down the time required to complete a single transaction, such an HTTP request.

You may find it helpful to compare the end-to-end time against other metrics, such as throughput, to see how increases in traffic impact the ability of the application to respond. For instance, if new connections double but the end-to-end time quadruples, you may need to consider adding additional servers. You can expand or collapse this pane by clicking the **+** or **–** icon beneath the health score, as shown.



From left to right, this pane displays the following timing information:

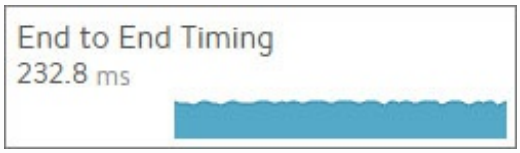
- **Client RTT:** Average network TCP latency between the client and Avi Vantage for all clients, both local (within the same datacenter) and remote (Internet). This metric indicates how long it takes to establish connections and return acknowledgements. Because this is TCP, this number will often be higher than an ICMP ping.
- **Server RTT:** Round-trip latency for SE-to-server traffic. An abnormally high server RTT may indicate either that the network is saturated or more likely that a server's TCP stack is overwhelmed and cannot quickly establish new connections.
- **App Response:** The time the server took to respond. This includes the time the server took to generate content, potentially fetch backend database queries, and begin transferring the response back to Avi Vantage. This time is calculated by subtracting the server RTT from the time of the first byte of a response from the server. If the application consists of multiple tiers (such as web, applications, and database), then the App Response represents the combined time before the server in the pool began responding. This metric is only available for a layer 7 virtual service.
- **Data Transfer:** Average time required for the server to transmit the requested file. This is calculated by measuring from the time the SE received the first byte of the server response until the client has received the last byte, which is measured as the when the last byte was sent from the SE plus one half of a client round trip time. This number may vary greatly depending on the size of objects requested and the latency of the server network. The larger the file, the more TCP round trip times are required due to ACKs, which are directly impacted by the client and server RTT. This metric is only used for a layer 7 virtual service.
- **Total Time:** Total time from when a client sent a request until they received the response.

This is the most important end-to-end timing number to watch, because it is the sum of the other four metrics. As long as the Total Time value is consistently low, the application is probably successfully serving traffic.

Metrics Tiles

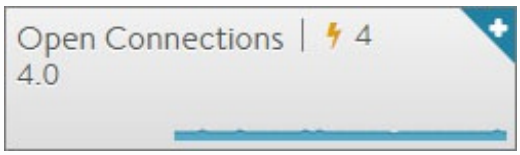
The following metrics are available for virtual services:

- **End to End Timing:** Displays the same information shown in the End-to-End Timing Pane in a chart form, with color-coding to show the client RTT, server RTT, data transfer, and app response times. Clicking the radio button for these items toggles the chart display of that item on and off. For non-HTTP virtual services, only client RTT and server RTT are shown.

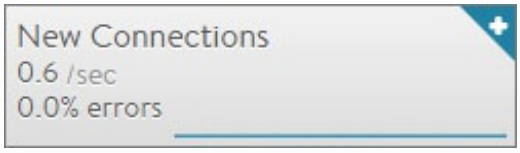


- **Throughput:** Total bandwidth passing through the virtual service, in Mbps. Hovering the mouse cursor over this graph displays the throughput in Mbps for the selected time. Throughput shown is measured as bytes transferred between the client and Service Engine, which does not include data transferred between Service Engine and the servers.

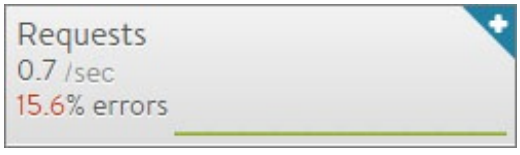
- **Open Connections:** Number of TCP client connections or UDP transactions that are currently in an open state. UDP transactions are counted even though they are technically connection-less. The number of open connections does not necessarily translate into the number of unique clients, as protocols such as HTTP 1.1 typically open six connections per client browser. If the Maximum Concurrent Connections setting has been set for a virtual service, a horizontal red bar superimposed over the chart pane will show the maximum number of connections allowed.



- **New Connections:** Number of established client TCP or UDP connections that were closed in the selected time period. The New Connections metric also includes a sub-metric for errors. Errors are defined as dropped, reset, or lossy connections. Errors can be customized via the Analytics profile assigned to the virtual service. These errors typically occur during the connection, but not necessarily during the first client packet. This is why the New Connections metric technically measures closed connections, rather than new connections. See this article for further description of this behavior.



- **Requests:** Number of responses to requests per second, which breaks down successful requests versus errors (such as 4xx or 5xx errors). Within the chart pane you

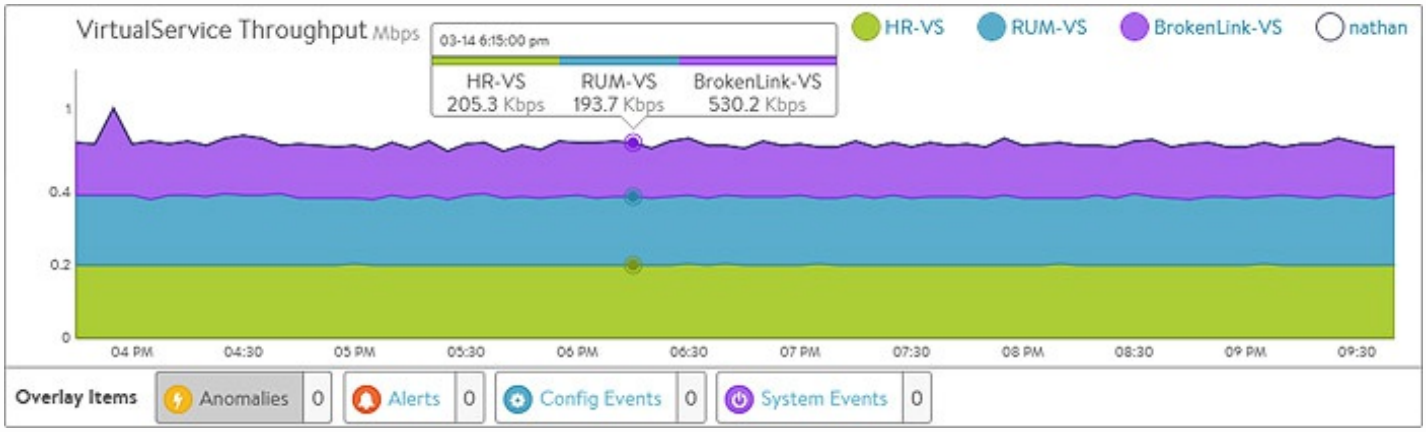


may toggle radio buttons to break down the errors by whether they are generated by the server or by Avi Vantage. For instance, if no servers are available in the pool, Avi Vantage will send out an HTTP 500 status code. The Client Logs section will provide more detail on specific errors. This metric is only available for HTTP virtual services.

Chart Pane

The main chart pane in the middle of the Analytics tab displays a detailed historical chart of the selected metric tile for the current virtual service.

- Hovering the mouse over any point in the chart will display the results for that selected time in a popup window.
- Clicking within the chart will freeze the popup at that point in time. This may be useful when the chart is scrolling as the display updates over time.
- Clicking again will unfreeze the highlighted point in time.

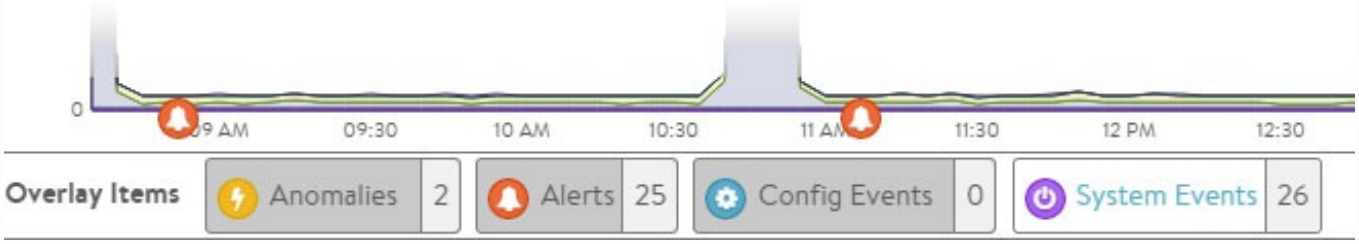


Many charts contain radio buttons in the top right that allow you to customize which data should be excluded from the chart. For example, if the End to End Timing chart is heavily skewed by one very large metric, then deselecting that metric by clearing the appropriate radio button will re-factor the chart based on the remaining metrics shown. This may change the value of the vertical Y-axis.

Some charts also contain overlay items, which will appear as color-coded icons along the bottom of the chart.

Overlays Pane

The overlays pane allows you to overlay icons signifying important events within the timeline of the chart pane. This feature helps correlate anomalies, alerts, configuration changes, or system events with changes in traffic patterns.



Within the overlays pane:

- Each overlay type displays the number of entries for the selected time period.
- Clicking an overlay button toggles that overlay's icons in the Chart pane. The button lists the number of instances (if any) of that event type within the selected time period.
- Selecting an overlay button displays the icon for the selected event type along the bottom of the chart pane. Multiple overlay icon types may overlap. Clicking the overlay type's icon in the chart pane will bring up additional data below the Overlay Items bar. The following overlay types are available:
    - **Anomalies:** Display anomalous traffic events, such as a spike in server response time, along with corresponding metrics collected during that time period.
    - **Alerts:** Display alerts, which are filtered system-level events that have been deemed important enough to notify an administrator.
    - **Config Events:** Display configuration events, which track configuration changes made to Avi Vantage by either an administrator or an automated process.
    - **System Events:** Display system events, which are raw data points or metrics of interest. System events can be noisy, and are best used as alerts that filter and classify these raw events by severity.

Anomalies Overlay

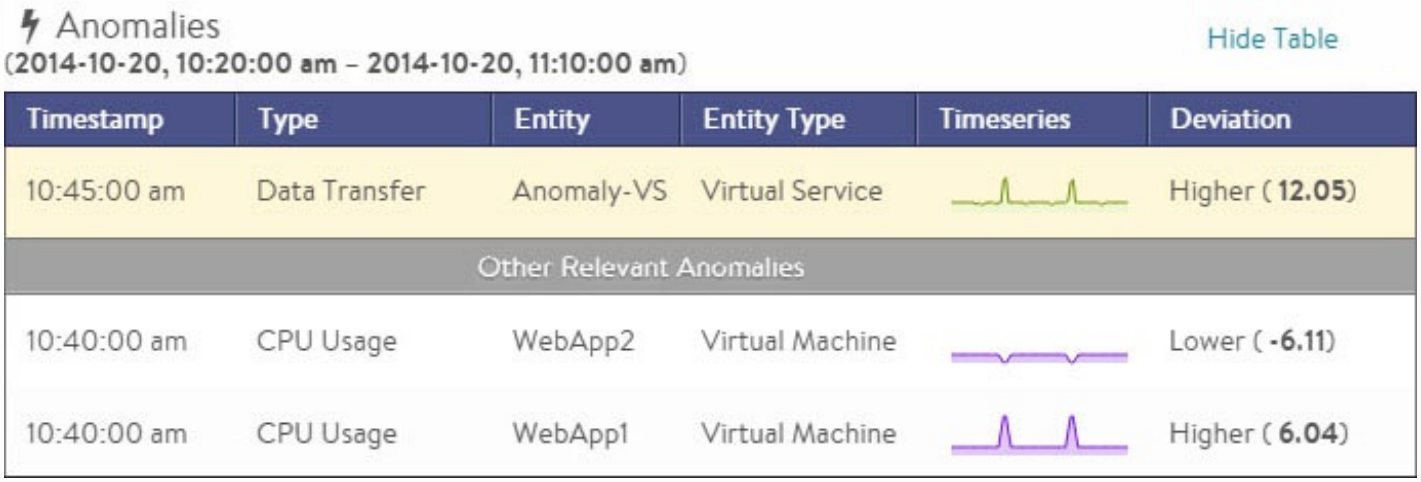
The Anomalies overlay displays periods during which traffic behavior was considered abnormal based on recent historical moving averages. Changing the time interval will provide greater granularity and potentially show more anomalies. Clicking the anomalies overlay button 

displays yellow Anomaly icons in the chart pane, which can scroll down to view additional information related to that anomaly. During times of anomalous traffic, Vantage records any metrics that have deviated from the norm, which may provide hints as to the root cause of the anomaly.

Note: An anomaly is defined as a metric that has a deviation of 4 sigma or greater across the

moving average of the chart.

Note: Anomalies are not recorded or displayed if the statistics interval (set from the Displaying pull-down menu) is set to Real Time.



These metrics are defined as follows:

- **Timestamp:** Date and time when the anomaly was detected. This may either span the full duration of the anomaly, or merely be near the same time window.
- **Type:** The specific metric deviating from the norm during the anomaly period. To be included, the metric deviation must be greater than 4 sigma. Numerous types of metrics, such as CPU utilization, bandwidth, or disk I/O may trigger anomalous events.
- **Entity:** Name of the specific object that is reporting this metric.
- **Entity Type:** Type of entity that caused the anomaly. This may be one of the following:
    - **Virtual Machine** (server). These metrics require Avi Vantage to be deployed in read access mode or write access mode to the virtualization orchestrator (for example, vCenter or OpenStack). In the example above, CPU utilization of the two servers was learned by querying vCenter.
    - **Virtual Service**
    - **Service Engine**
- **Time Series:** Thumbnail historical graph for the selected metric, including the most current value for the metric which will be data on the far right. Moving the mouse over the chart pane will show the value of the metric for the selected time. Use this to compare the normal, current, and anomaly time periods.
- **Deviation**: Change or deviation from the moving average, either higher or lower. The time

window for the moving average depends on the time series selected for the Analytics tab.

Alerts Overlay

The alerts overlay displays the results of any events that meet the filtering criteria defined via Operations > Alerts. Alerts notify administrators about important information or changes to a site that may require immediate attention.

Alerts may be transitory, as they will expire after a defined period of time. For instance, Vantage may generate an alert if a server is down and then allow that alert to expire after a specified time period once the server comes back online. The original event remains available for later troubleshooting purposes.

Clicking the alerts icon 

in the overlay items bar displays any red alerts icons in the chart pane. Selecting one of these chart alerts will bring up additional information below the overlay items bar, which will show the following information:

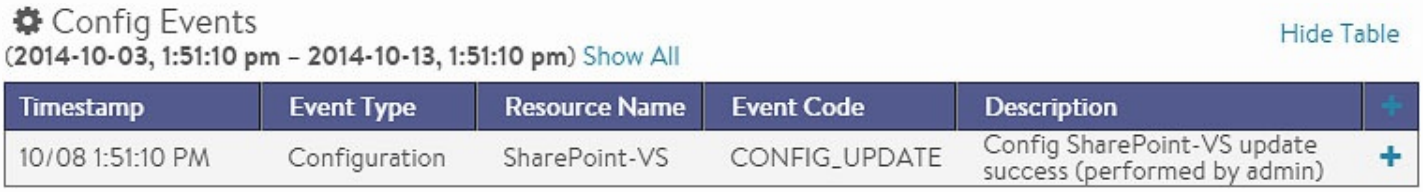
- **Timestamp**: Date and time when the alert occurred.
- **Resource Name**: Name of the object that is reporting the alert.
- **Level**: Severity of the alert. You can use the priority level to determine whether additional notifications should occur, such as sending an email to administrators or sending a log to Syslog servers. The level may be one of the following:
  - **High**: red
  - **Medium**: yellow
  - **Low**: blue
- **Summary**: Brief description of the event.
- **Actions**:

- **Dismiss:** Dismissing the alert removes it from the list and also removes the corresponding alert icon from the chart pane. Dismissing an alert here is the same as dismissing it via the bell icon at the top of the screen next to the health score, or dismissing it via the Alerts tab.
- **Edit:** Opens the alert filter configuration, which can be modified to make Vantage more or less sensitive to generating new alerts.

- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for an alert opens and closes a sub-table showing more detail about the alert. This will typically show the original events that triggered the alert.

Configuration Events Overlay

The configuration events overlay displays configuration events. Configuration events typically involve changing the Vantage configuration by adding, deleting, or modifying a resource such as a pool, virtual service, or SE, or an object related to the object being inspected. For example, if a virtual service's traffic drops off at precisely 10:00 a.m., and at that time an administrator made a change to the virtual service's security settings, there is a good chance the cause of the change in traffic can be attributed the configuration change.



Clicking the configuration events icon 

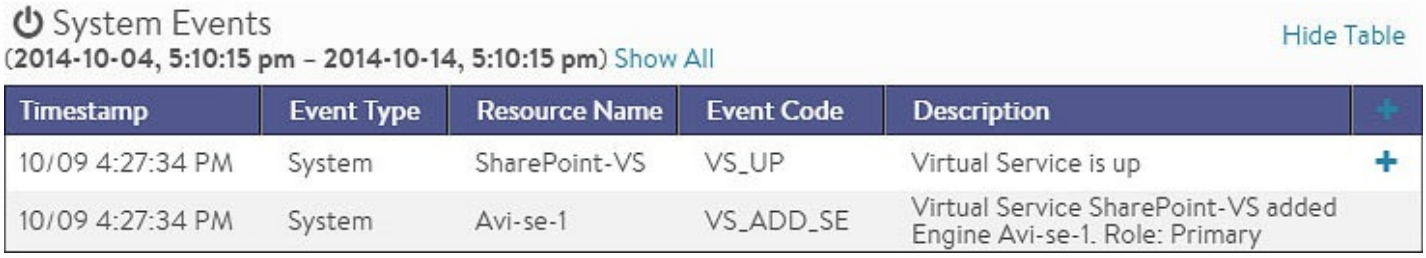
in the overlay items bar displays any blue configuration event icons in the chart pane. Selecting one of these chart alerts will bring up additional information below the overlay items bar, which will show the following information:

- **Timestamp**: Date and time when the configuration change occurred.
- **Event Type**: Always scoped to *Configuration* event types.
- **Resource Name**: Name of the object that has been modified.
- **Event Code**: Type of change:
  - CONFIG_CREATE
  - CONFIG_UPDATE
  - CONFIG_DELETE

- **Description**: Brief description of the event.
- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for a configuration event either expands or contracts a sub-table showing more detail about the event. When expanded, this shows a difference comparison of the previous configuration versus the new configuration, as follows:
  - Green: shows additions.
  - Red: shows deletions.
  - Yellow: shows changes.

System Events Overlay

This overlay displays system events relevant to the current object, such as a server changing status from up to down or the health score of a virtual service changing from 50 to 100.



Clicking the system events icon 

in the overlay items bar displays any purple system event icons in the chart pane. Select a system event icon in the chart pane to bring up more information below the overlay items bar.

- **Timestamp**: Date and time when the system even occurred.
- **Event Type**: This will always be system.
- **Resource Name**: Name of the object that triggered the event.
- **Event Code**: High-level definition of the event, such as VS_Health_Change or VS_Up.
- **Description**: Brief description of the system event.
- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for a System Event expands or contracts that system event to show more information.

Virtual Service Logs

Virtual services and pools are able to log client-to-application interactions for TCP connections and HTTP requests/responses. These logs can be indexed, viewed, and filtered locally within the Avi Controller. Logs can be useful for troubleshooting and surfacing insights about the end-user experience and success of the application.

Enabling Logs

See the Analytics tab of the Create Virtual Service popup for configuring, enabling, filtering, and/or disabling client logs.

Significant Logs

Avi Vantage automatically logs common network and application errors under the umbrella of *significant logs*. These significant logs may also include entries for lesser issues, such as transactions that completed successfully but took an abnormally long time.

Errors may include any of the following:

- HTTP errors, such as server or Vantage-originated 4xx and 5xx errors
- Network errors, such as aborted connections, abnormal latency, or out of order packets.
- See Log Events for a list of error events that may trigger a significant Log.

Errors can be omitted from the significant logs list by editing the analytics profile used by the virtual service.

Full Client Logs

In addition to significant logs, a virtual service may be configured to log all client connections or HTTP requests. The Full Client Logs option includes any significant logs, custom full log filters, and any logs generated by custom policies or DataScripts. By default, a new virtual service is configured to provide full client Logs for the first 30 minutes, then drop down to a reduced logging level by capturing significant logs only. From the Analytics tab, full client logs may be enabled for the virtual service, either temporarily or permanently.

Full client log filters may also be specified for IP addresses or URIs, which is recommended when capturing important information from busy production systems. An additional level of logging is provided by enabling the All Headers option in a client log filter. This option will capture all

headers from the client and server within the logs. Keep in mind this may have significant impact on the size of the logs, as some applications send as much as 30 k within a single header. Even so, the All Headers option is very useful for quick troubleshooting to see what each side of the connection is sending and receiving.

When logging, Avi Vantage prioritizes significant logs by proactively sending them to the Controller for indexing. Non-significant logs, such as normal HTTP request logs, will be pulled from the SEs and indexed by the Controller ***only*** when an administrator attempts to view full client logs for the virtual service or pool. This may take anywhere from a few seconds to hours to process. Logs will be viewable while the indexing process is performed in the background. This time may depend on network latency from the SEs to the Controllers, the volume of logs, and the hardware used by the Controller for performing the resource-intensive task of indexing the data.

Rotation Out of Unrequested Logs

Capturing all logs can consume significant resources. For this reason, logs that are unrequested are rotated out of the SE's storage after a period of time. The allocation for raw log storage on an SE is variable, but a minimally configured SE with 2 vCPUs, 2 GB memory, and 10 GB storage can store about 8 million logs. Adding more resources (CPU, memory, and especially disk space) to the Controllers and SEs will extend both the volume of logs per second and the length of time that Vantage can store the logs.

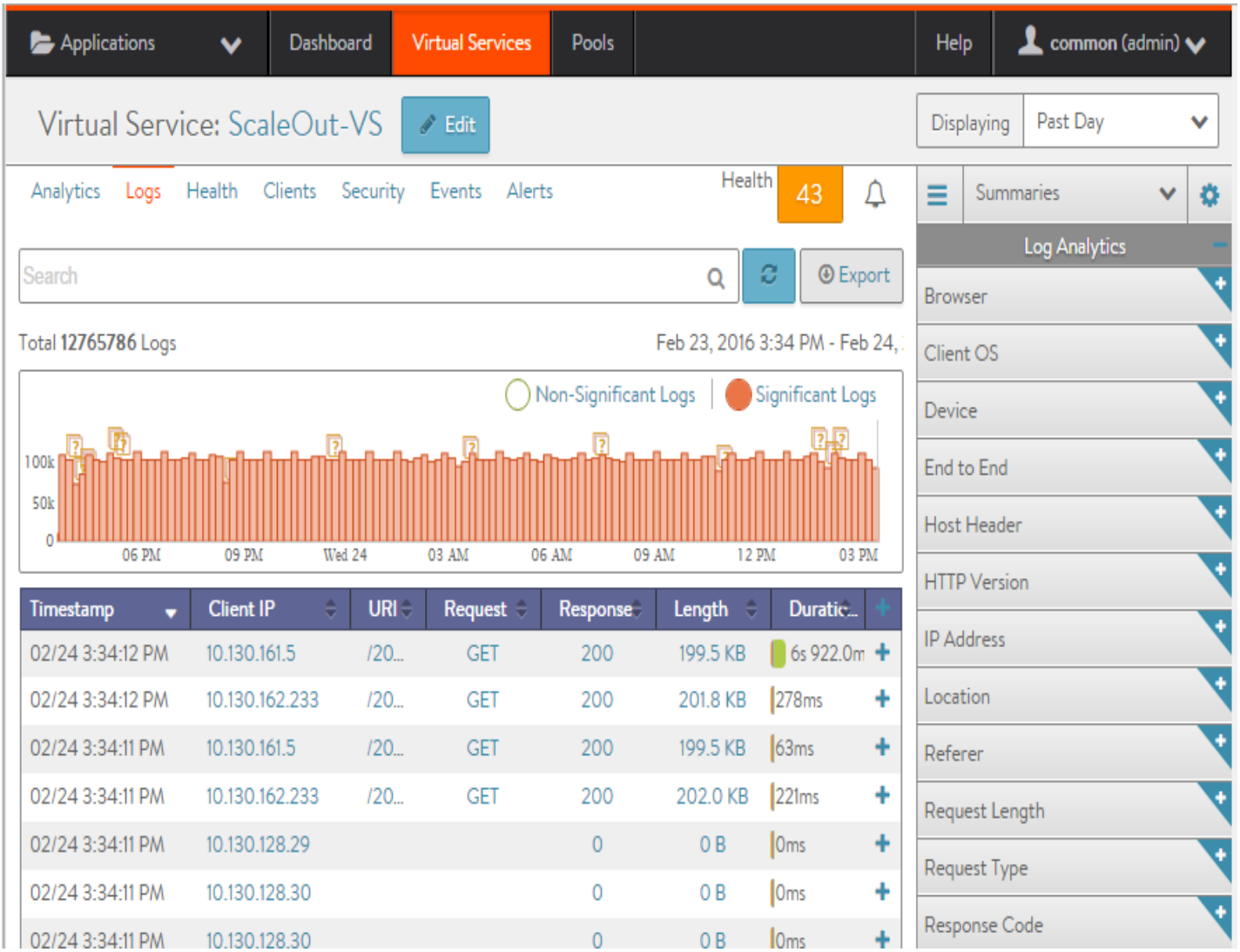
Logging During Heavy Load

In a very busy system with high volumes of requests per second, Vantage may temporarily degrade to capturing only a sampling of requests. If the system is set up with redundant Controllers, the task of indexing logs for multiple virtual services is automatically shared across the Controllers to better utilize resources.

Log Navigation

To access the client logs for a virtual service:

1. Navigate to Applications > Virtual Services.
2. Click on the virtual service name.
3. Click on Logs.

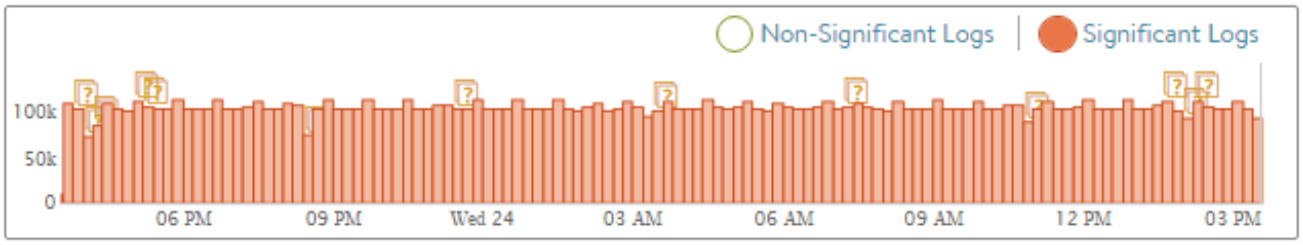


The top of the Client Logs tab displays the following items:

- **Search:**  The Search field allows you to filter the log entries using free-form text, keywords and/or a formal search syntax.
- **Refresh:**  Clicking Refresh updates the logs for the currently-selected time period.
- **Number and Range:**

  Total number of log items being displayed. The date/time range of those logs appear beneath the Search field on the left. The number will reflect how many logs passed the filter, which may be lower than the total number of logs.
- **Clear Selected:**  If one or more filters have been added to the Search field, click the gray X icon on the right side of the search bar to remove those filters. To clear individual filters, click the white on blue X for that filter.
- **Histogram:**



The histogram shows the number of client log entries over the selected time period. The X-axis is time, while the Y-axis is the number of logs during that bar's period of time.

- Green bars show All Logs, or successful transactions.
- Red bars show errors, or Significant.
- Orange question marks above the bars may appear, indicated potentially missing logs. This may happen for a number of reasons, most probably due to the SE dropping into a log sampling mode due to high volumes of logs for its allocated hardware resources.

Hovering the cursor over a Histogram bar displays the number of entries represented by that bar.

Click and drag inside the histogram to refine the date/time period which further filters the logs shown. Clicking the blue magnifying glass icon in the drawn box will zoom the Histogram to that view. This expands the drilled in time to zoom in to the width of the Histogram, and also changes the Displaying Time pull-down menu to Custom. To return to the previously selected time period, use the Displaying pull-down menu.

The Non-Significant and Significant options display *all* logs or only significant logs, respectively.

- **Export:** The logs may be downloaded directly from Vantage via the Export button to the right of the search bar. The logs are in comma-separated value (CSV) format. The downloaded file will include only the logs that match any custom filters, the Non-Significant Logs versus Significant Logs option, and the selected time period. This may include only the logs from the displayed page or all logs that match the filter, up to a maximum of 10,000 logs.

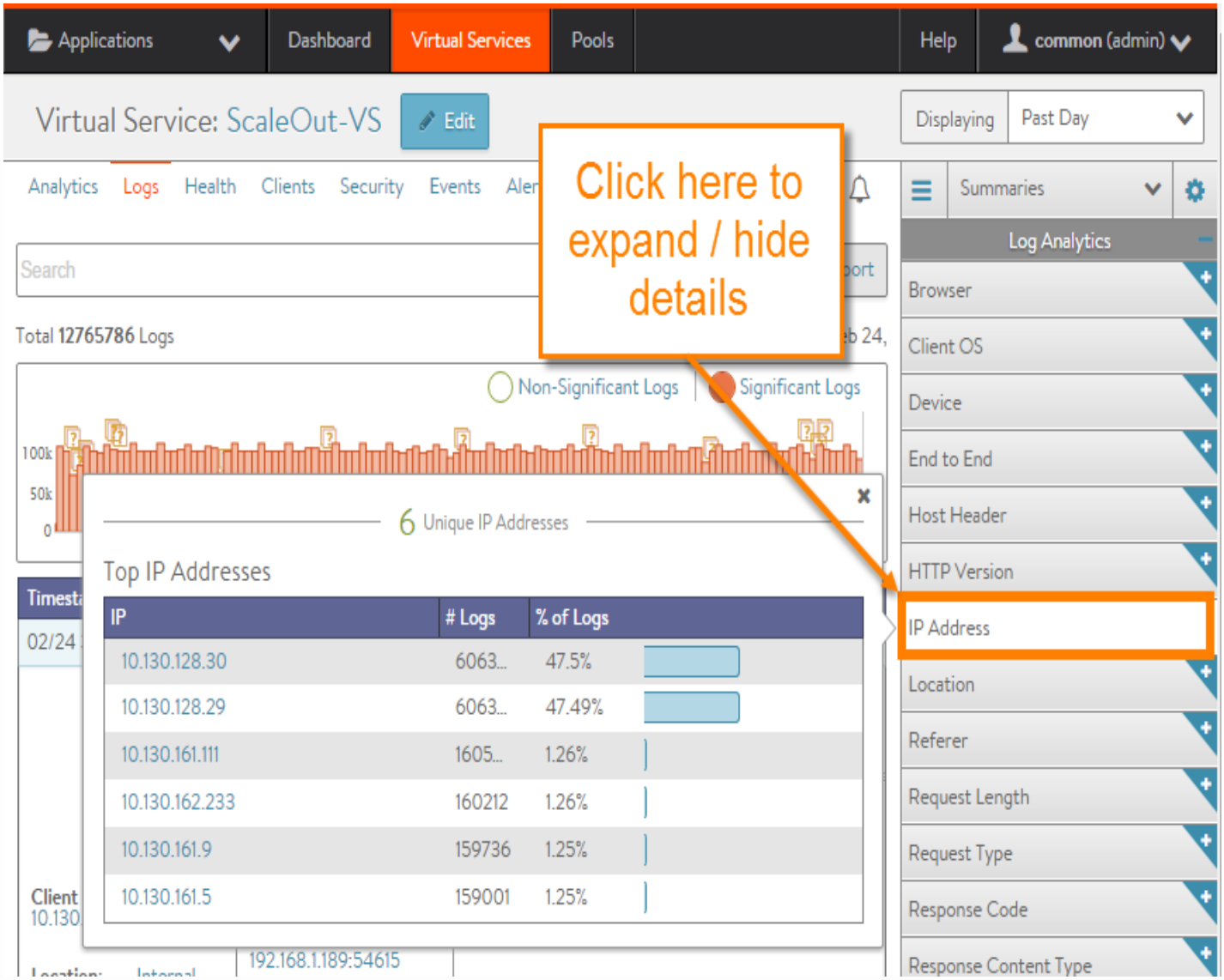
The primary table on the Logs tab displays the list of individual log entries. The following information appears for each log entry:

- **Timestamp:** Date and time the event occurred.
- **Client IP:** The source IP address of the client to which the log entry applies.
- **URI:** For HTTP enabled virtual services, the URI path of the request, such as */index.html*.
- **Request:** Type of HTTP method for the request, such as GET, POST, or HEAD.

- **Response:** For HTTP, the response code will be of type 1xx, 2xx, 3xx, 4xx, or 5xx. This is usually a reflection of the response sent by the server, but it may instead be a response created by Avi Vantage. Expand the log to determine whether the server or Avi Vantage generated the response.
- **Length:** Size of the response, in Kilobytes.
- **Duration:** Bar showing the total transaction time, which can be used to compare this request to others based on the length of the bar. Hovering the mouse pointer over the bar shows a breakdown of time that makes up the duration of the request. This is a miniature version of the End to End Timing chart that appears when the log is expanded.
- **Timeline:** This view creates a waterfall display of the transactions relative to each other. This is most useful when following a single client interacting with the site, because the timeline makes it easy to see requests that are blocking or congesting the client from loading the page.
- **More Info icon:** Clicking the + icon in the table header expands or collapses a view into logs. This can be done for all logs on the page, or individual logs.

Log Analytics

The Log Analytics area displays a series of prebuilt filters that summarize the client logs in a popup according to the selected summary filter. For example, clicking the IP Address summary tile will display the most active client IP addresses that have requested the virtual service within the log display period.



The Log Summaries reflect the currently applied filters, including the displayed log period and the Non-Significant/Significant setting.

Client information is learned from IP addresses or the client's HTTP User-Agent string.

Adding a search filter such as client_location="US" will filter the summaries to reflect only the results for clients coming from an IP address within the United States.

Log Analytics Tiles

- **Browser:** Client browser, such as Firefox, IE, or Safari.
- **Client OS:** Client operating system, such as iOS, Windows 7, or Mac OS X.
- **Device:** Type of client device, such as Computer, iPhone, Tablet, or Other.
- **End to End:** View of the client-to-server transaction, similar to the End to End Timing pane in the Analytics tab of the virtual service Details page.

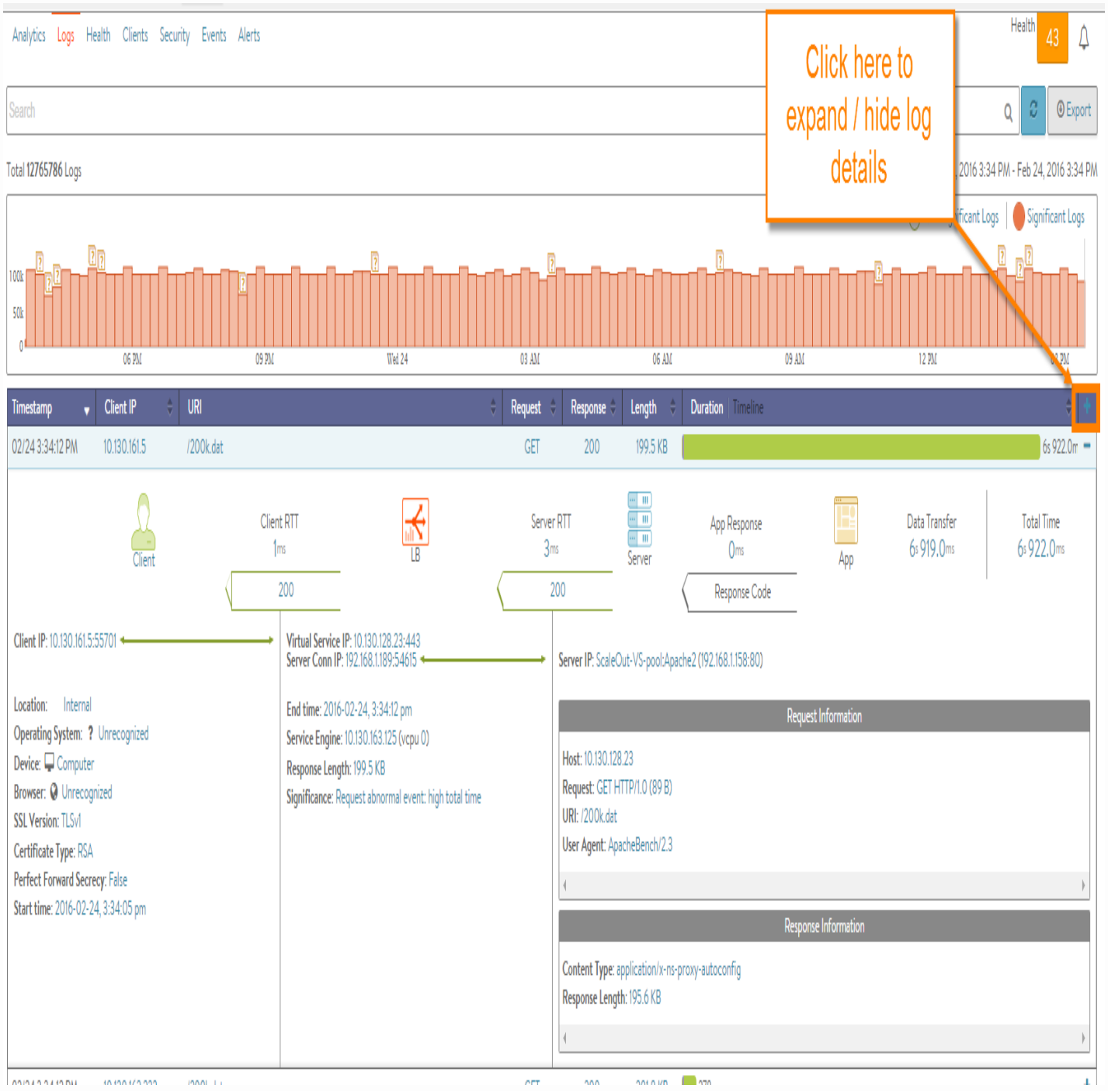
- **Host Header:** The most common HTTP Hosts (domain names) requested. This is most applicable in a forward proxy or virtual hosting.
- **HTTP Version:** The version of HTTP negotiated with the client. This may be 1.0, 1.1, or SPDY 3.1.
- **IP Address:** Number of unique client IP addresses, and the list of the top addresses in the logs.
- **Location:** Geographic location of clients. This may be a two letter country code or a custom location based on a match with an IP group.
- **Referer:** Top websites that have referred traffic, based on the HTTP Referer header.
- **Request Length:** Size of the HTTP requests, in bytes. HEAD requests are typically very small, GETs tend to be a bit larger, and POSTs can be extremely large.
- **Request Type:** A breakdown of the various HTTP request methods, such as GET, POST, HEAD, CONNECT, PUT, and so on.
- **Response Code:** The HTTP responses sent from Avi Vantage or the servers. This could be 200, 301, 304, 404, 503, and so on. Adding the IP address of a suspect server to the filter can give you a better view of the typical responses that the selected server is sending out.
- **Response Content Type:** Types of content requested by clients.
- **Response Length:** Size distribution of HTTP responses sent to clients. This gives a good indication of the average HTTP response size.
- **Server IP Address:** The list of the top servers. Add a filter such as resp_code>=500 to see which servers are responding with ‘server busy’ errors.
- **Significance:** A log is classified as Significant if it has an entry in the Significant log field. This is generally due to an error, such as client or server reset, out of order packets, HTTP 4xx or 5xx.
- **SSL:** Insights about SSL traffic that is terminated at the virtual service.
    - The SSL Version negotiated, which could be TLS 1.0, TLS 1.1, or TLS 1.2.
    - Perfect Forward Secrecy means that clients were able to negotiate a cipher that cannot be decrypted after the session ends using hijacked SSL keys.
    - The Authentication Protocol is either RSA or Elliptic Curve (EC).
    - The Encryption Protocol is the cipher negotiated for handling bulk encryption of data,

such as AES256 or AES128.

- **URL Path:** The most commonly requested HTTP paths, such as /index.html. Try adding a filter for 404 to see the most common broken links on your website or a filter for response time greater than 500 to see which objects are taking the longest to generate by the servers.
- **User ID:** Most common user IDs.

Expanded Logs

Clicking the **+** plus icon on the right of the logs table expands an individual log. This provides an in-depth view of the specific connection log or the HTTP request and response log.



- **End to End Timing:** The bar is similar to the Analytics tab of the virtual service Details page, though it also contains arrows indicating the HTTP response code. This data is specific to this single connection or HTTP request, whereas the Analytics tab for the virtual service shows an aggregate of all connections or requests. If the arrow under Server RTT is a 0, then no response was received from the server. This may be due to an error such as a timed out server response, or because the request was served by Avi Vantage, such as via caching or a policy.
- **IP Addresses:** Under End to End Timing, IP addresses and service ports indicate the client source address and port being used to initiate a transaction to the virtual service IP address and destination service port. The second address under the Avi icon is the Avi Vantage source NAT (SNAT) address and source port that is used for communicating with the destination server's pool, name, IP address and port.
- **Client data:** The first column on the bottom left displays information regarding the client.
    - **Client IP:** The source IP address and service port of the client.
    - **Location:** The country of origin for the IP address or Internal for private IP addresses. This field may also show custom IP group names.
    - **Operating System:** The OS of the connecting device. HTTP only.
    - **Device:** The type of connecting device, such as computer, table, or phone. HTTP only.
    - **Browser:** The web browser of the connecting device. HTTP only.
    - **SSL Version:** The negotiated version, such as SSL 3.0, TLS 1.0, TLS 1.1, or TLS 1.2. SSL terminated HTTP traffic only.
    - **Certificate Type:** RSA or Elliptic Curve (EC) certificate used for the connection. SSL terminated HTTP traffic only.
    - **Perfect Forward Secrecy:** Did the client negotiate a cipher which protects the connection from later decryption via hijacked keys. SSL terminated HTTP traffic only.
    - **Start Time:** The time the connection was established or the request was received.
- **Avi data:** The following information appears under the Avi icon in the middle column:
    - **Virtual Service IP:** The listening virtual service's IP and port.
    - **Server Conn IP:** The source IP address and port used as the source NAT address on the server side of the connection.

- **End time:** When the log was generated; this normally occurs when the request or connection was completed; however, logs may be generated for currently active, long-lived connections. Logs generated during an open connection will be updated periodically or when the connection closes.
- **Service Engine:** The SE and corresponding vCPU that was used to process the request or connection.
- **Response Length:** The size of the response, such as HTTP payload plus headers returned by Avi Vantage to the client. This size may be different from the server Response Length in the server column due to SSL padding, Javascript insertion (when Client Insights is set to Active), compression, TCP maximum segment size differences, or a number of other features.
- The following fields appear only if applicable:
  - **Cache Hit:** This is true if the HTTP request was served by the Avi Vantage cache. This field will not be shown if caching is disabled.
  - **Compression:** If Avi Vantage compressed the response content, this will show the percent by which the content was able to be compressed.
  - **Policy Rule:** If a policy has been applied to the virtual service, any rules that were executed will be displayed. If the rule was created with the log checkbox enabled, the log will be generated even if the virtual service does not have full client logs enabled on the Analytics tab for the virtual service. These logs will still require Non-Significant Logs to be selected in order to be displayed (unless they qualify as Significant Logs, of course).
  - **Significance:** If the connection or request is determined to be an error, it will be marked as Significant. This field provides a description of the issue (such as client terminated the connection or server returned 500 error).

- **Server and App data:** The third column on the bottom of an expanded log provides information about the connection or request and response.
  - **Server IP:** Pool name, server name, and the server IP address and port.
  - **Host:** The HTTP Host header, such as www.avinetworks.com or 10.1.1.10.
  - **Request:** The HTTP method (such as GET), version (such as HTTP/1.1), and size of the request (such as 2 Kb).
  - **URI:** The HTTP path and query of the client request.

- **User Agent:** The raw client HTTP User-Agent header (such as Mozilla/5.0, AppleWebKit/533, and so on).
- **Content Type:** The HTML, images, Javascript, and so on returned to the client.
- **Response Length:** The size of the HTTP header plus content returned from the server to Vantage. (This may be different from the size of the response length from Vantage to the client due to compression, inserting JavaScript, or other acceleration that may alter the content size before it is sent to the client.)

- **View All Headers:** A View All Headers option may exist within the expanded logs. This button will further expand the log display to show additional information for the transaction. View All Headers may exist due to one or both of the following:
  - **All Headers:** On the Analytics tab for the virtual service, create a new filter with the All Headers option selected. This will cause Vantage to record all client request and server response headers. Custom headers, cache control, and other useful troubleshooting may be done by viewing full headers. Headers may be quite large, which could come at the cost of a significant resource hit to the SEs creating the logs and the Controller storing the larger logs. Recommendation is to turn this feature on selectively, such as for specific clients or for a shorter time duration.
  - **DataScript Errors:** Many DataScript errors are caught when attempting to save a new script. However, there are many scenarios when the script may fail when executing. When this happens, an error will be created in the logs, visible under View All Headers. The error and stack trace may be included to help determine the cause of the error.

Searching Logs

The Search field above the list of client log entries filters the logs according to your specified search terms. You may use either arbitrary search strings or a specific search syntax. For example, entering “mobile” will filter the logs to all entries that include this string anywhere within the log. **Search strings are *not* case sensitive.**

To use the formal search syntax, the search filter may be typed manually, or clicking any blue text within a log entry will generate the filter. For example, clicking Client IP creates “filterclient_ip=A.B.C.D” where A.B.C.D is the client’s IP address. In this example, the logs will be filtered to show only clients with that specific IP address.

When typing a search directly into the Search field, contextual help will show the available options. For instance, typing “client_ip” will show the appropriate operands (described below)

such as “ = “.

Example: The filter client_ip= will show the most commonly seen IP addresses, including the number of logs generated by these addresses and the percentage of logs from this address, in this format: 10.30.4.31 - 15924 - (34.7%)

When using the search syntax:

- Be aware of whether the log display filter is set to Significant Logs or Non-Significant Logs.
- Multiple filters may be selected or created to further refine a search. All filters must be true for the log to match.
- Search criteria for strings must be enclosed in quotation marks, such as client_country=”US” or”Connection setup failed”.
- The filter can include any combination of informal strings and formal search syntax, such as: iphone client_ip^=”10.30.”
- Searches may be saved and reused later.
- The following operators allow more granular searches for strings, numbers, or IP addresses:
  - **Greater than:** >
  - **Greater than or equal:** >=
  - **Less than:** <
  - **Less than or equal:** <=
  - **Not equal:** !=
  - **Equals:** =
  - **Contains:** ~=
  - **Does not contain:** !~=
  - **Starts with:** ^=
  - **Ends with:** $=

Significant Log Events

Avi Vantage records the following types of significant log events for clients (grouped alphabetically by high-level description).

- **Avi rule failed to execute:** The Avi Vantage policy or DataScript failed to execute. For DataScript failures, the expanded log may have a “View All Headers” button, which brings up a modal containing the DataScript error and stacktrace.
- **Connection abnormal event:** The TCP connection established successfully, but there are a high number of retransmitted packets due to timeouts from Avi Vantage to the client. This could be due to lossy network or a slow client.
- **Connection abnormal event:** The TCP connection established successfully, but there were a high number of retransmitted packets due to timeouts from Avi Vantage to the server. This could be due to lossy network or a slow server.
- **Connection abnormal event:** The TCP connection completed gracefully, but there were instances when the TCP window advertised by the client depleted to zero. This means that Avi Vantage could not send any more packets until the client opened the TCP window. This is typically due to a slow client.
- **Connection abnormal event:** The TCP connection completed gracefully, but there were instances when the TCP window advertised by the server depleted to zero. This means that Avi Vantage could not send any more packets until the server opened the TCP window. This is typically due to a slow server.
- **Connection abnormal event:** The TCP round trip time between Avi Vantage and the client is greater than 1 second.
- **Connection abnormal event:** The TCP round trip time between Avi Vantage and the server is greater than 500ms.
- **Connection abnormal event:** The TCP connection completed gracefully, but there were a high number of retransmitted packets from Avi Vantage to the client. This may be due to a lossy or congested network.
- **Connection abnormal event:** The TCP connection completed gracefully, but there were a high number of retransmitted packets from Avi Vantage to the server. This may be due to a lossy or congested network.
- **Connection abnormal event:** The TCP connection completed gracefully, but Avi Vantage received a high number of out of order packets from the client. This is generally due to lossy or congested network or drops by Avi Vantage.
- **Connection abnormal event:** The TCP connection completed gracefully, but Avi Vantage received a high number of out of order packets from the client. This is generally due to lossy or congested network or drops by Avi Vantage.

- **Connection closed abnormally:** Client abnormally terminated the connection by sending a TCP RST.
- **Connection closed abnormally:** Server abnormally terminated the connection by sending a TCP RST.
- **Connection closed abnormally:** The established TCP connection with the client was closed due to timeout. Despite multiple retry attempts, the client stopped responding to Avi Vantage's packets.
- **Connection closed abnormally:** The established TCP connection with the server was closed due to timeout. Despite multiple retry attempts, the server stopped responding to Avi Vantage's packets.
- **Connection closed abnormally:** The connection timed out waiting for the client to send HTTP request.
- **Connection closed abnormally:** The connection was prematurely closed by the client.
- **Connection closed abnormally:** The connection dropped due to SSL handshake failure.
- **Connection closed abnormally:** The connection was closed due to internal error.
- **Connection closed abnormally:** Validation of the server's x509 certificate failed.
- **Connection closed abnormally:** The connection was dropped due to server side SSL handshake failure.
- **Connection closed abnormally:** The connection was deleted due to an Avi Vantage configuration update.
- **Connection closed abnormally:** The established TCP connection was closed due to the TCP profile's idle timeout triggered.
- **Connection setup failed:** The connection setup was aborted by the client, which sent a RST during the TCP three way handshake.
- **Connection setup failed:** The connection setup was refused by the server, which sent a RST during the TCP three way handshake.
- **Connection setup failed:** TCP three way handshake with the client timed out despite multiple retries.
- **Connection setup failed:** TCP three way handshake with the server timed out despite multiple retries.

- **Connection setup failed:** Avi Vantage was unable to complete the TCP three way handshake with the client due to an internal Avi Vantage error. This may be due to lack of memory or other resources.
- **Connection setup failed:** Avi Vantage was unable to complete the TCP three way handshake with the server due to an internal Avi Vantage error. This may be due to lack of memory or other resources.
- **Connection setup failed:** The connection was dropped due to bad packets received from client during TCP three way handshake.
- **Request abnormal event:** The response time was greater than the configured threshold.
- **Request abnormal event:** Server response time greater the configured threshold.
- **Request abnormal event:** The server this connection was persisted to is no longer a valid destination.
- **Request abnormal event:** Client sent request with HTTP Protocol version 0.9.
- **Request ended abnormally:** A 4xx ‘client error’ HTTP response code was sent to the client.
- **Request ended abnormally:** A 5xx ‘server error’ HTTP response code was sent to the client.
- **Request ended abnormally:** The connection attempt was reset due to no servers available. This may be because the server connection limits setting was reached, or all servers are in a down state.

Virtual Service Health Monitoring

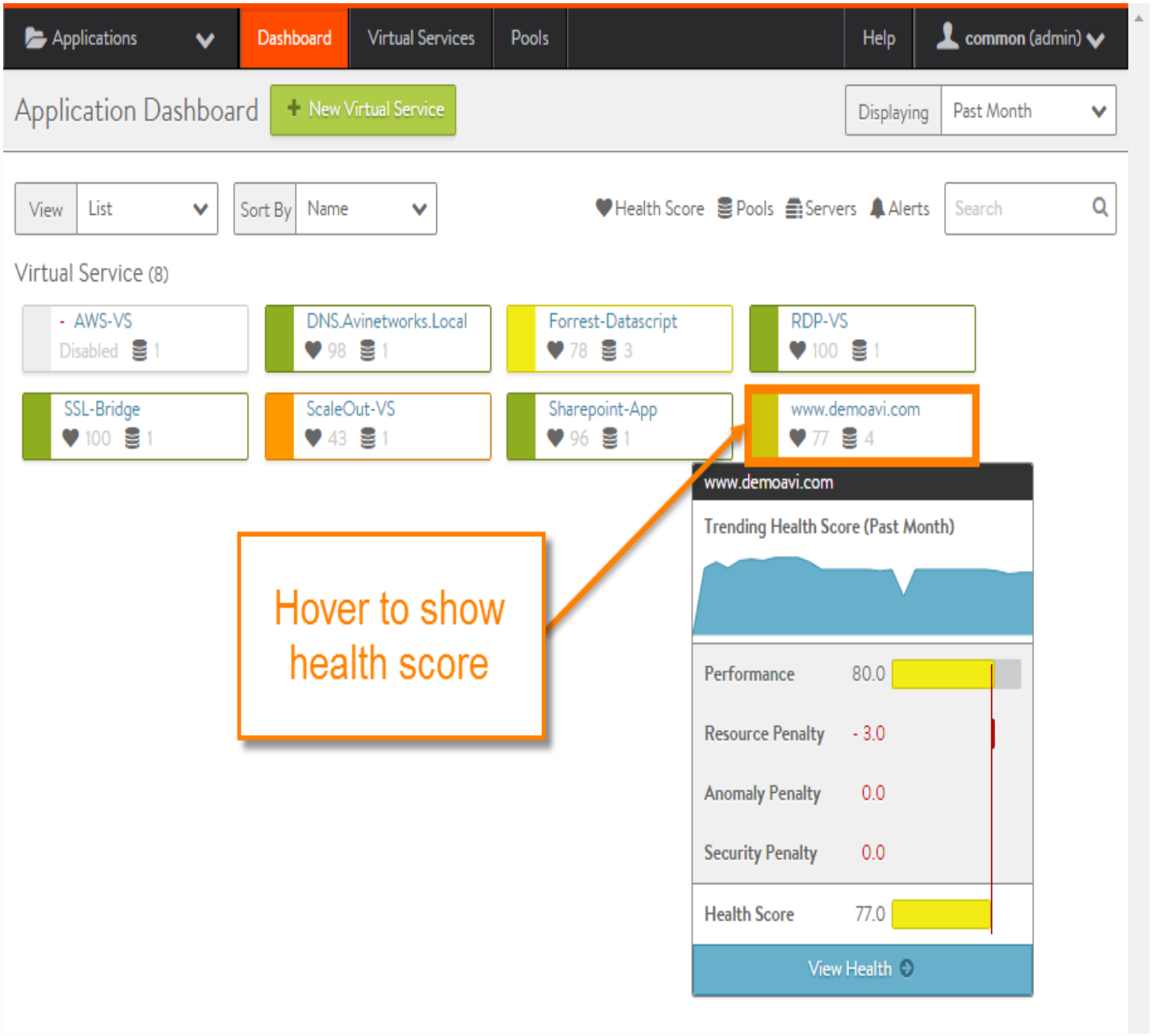
The Health tab presents a detailed breakdown of health score information for a virtual service. The health score is a pervasive concept throughout the Avi Vantage web interface.

Displaying Virtual Service Health

Health information for a virtual service can be viewed in summary form by hovering over the virtual service icon. Detailed health information is available on the Health tab for the virtual service.

Summary Health Scores

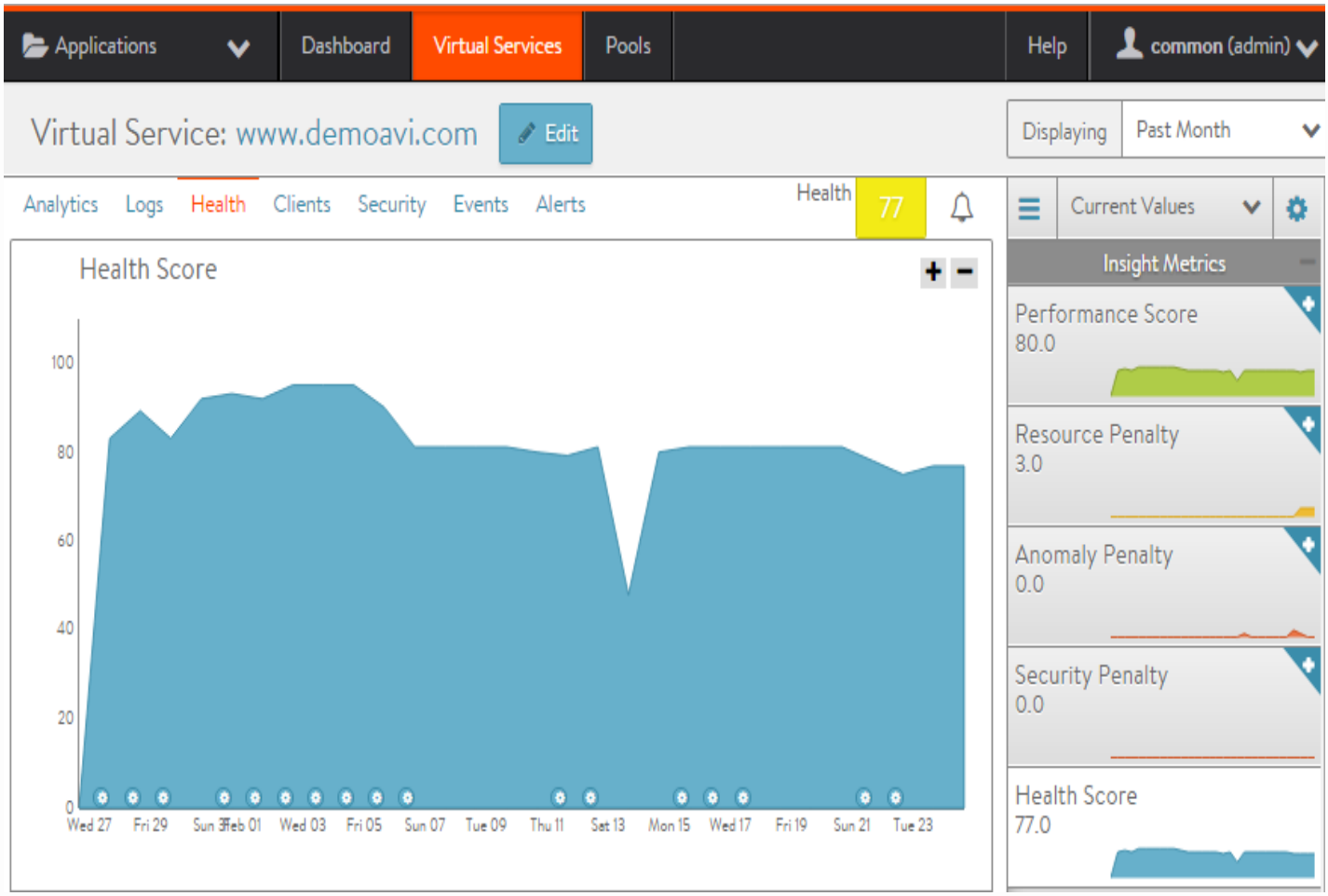
On the application dashboard, hovering over a virtual service icon displays the health score for the virtual service:



Health Details

Navigating to the Health tab for the virtual service provides a more robust view into the components that comprise the score, along with their historical trends.

1. Navigate to Application > Virtual Services.
2. Click on the virtual service name.
3. Click on Health.



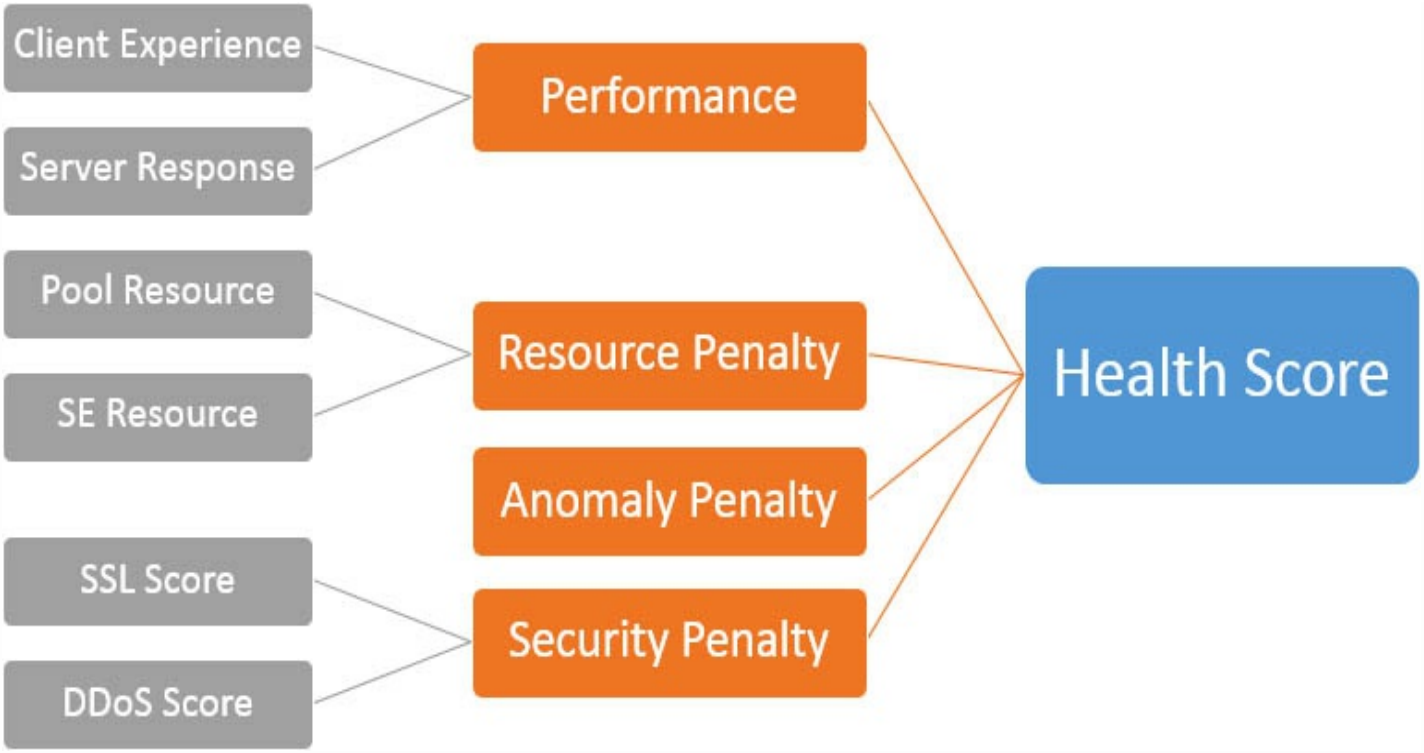
Health Score Codes

The health score of a virtual service is a composite of several scores. The final score denotes both the responsiveness of the virtual service ***and*** any vulnerabilities.

Vantage uses health score colors to show the health range of the object:

- **Green:** (health score 85 to 100).
- **Yellow:** (health score 65 to 84)
- **Orange:** (health score 1 to 64)
- **Red:** (health score '!', or the server is down)
- **Gray:** Dash '-'; object is not available to use, such as disabled.

Health Score Composition



- **Performance:** Performance is comprised of end user's experience and server response. The performance score is a positive number from 0 (down) to 100 (ideal).
    - **Client Experience:** Based on the end to end timing, this determines if the client sees the site as fast or slow. The Analytics profile can be used to determine the window of time for a satisfactory response.
    - **Server Response:** This will lower the score if the server is sending back errors such as TCP resets or HTTP 500 (busy).
- **Resource Penalty:** Any penalty assessed because of resource availability issues is assigned a score, which is then subtracted from the performance score. A penalty score of 0 is ideal, meaning there are no obvious resource constraints on Vantage or virtualization orchestrator connected servers.
    - **Pool Resource:** If the Controller is able to gather resource metrics from servers, such as via read or write access modes to vCenter, then servers saturating their CPU, memory, or disk IO may incur a penalty to the pool. The resource utilization is based on the average of a five minute polling interval. Thresholds may be determined within the Analytics Profile.
    - **Service Engine Resource:** If an SE is experiencing high CPU or memory utilization for an average five minute period, it will result in an SE penalty.
- **Anomaly Penalty:** *Anomaly* is defined as traffic that is outside the statistical norm for the time period. The algorithm takes into account seasonality for the virtual service. While

anomalies in traffic are not necessarily bad, they may denote potential risk.

- **Security Penalty:** The security penalty shows risks due to configuration settings that leave a site vulnerable, as well as vulnerability to attacks such as DDoS.
    - **SSL Score:** Configuration of SSL settings for an SSL-enabled virtual service will affect this score. This could be due to settings within the SSL/TLS profile (such as enabling an insecure cipher), due to the associated certificate (which may be expired or self-signed), or due to other SSL settings (such as not enabling HSTS in the HTTP profile).
    - **DDoS Score:** If the virtual service is experiencing a DoS attack, it will increase the security penalty depending on the severity of the attack.
- **Health Score:** The final health score for the virtual service equals the performance score minus the various penalty scores. The score may change depending on the window of time selected.

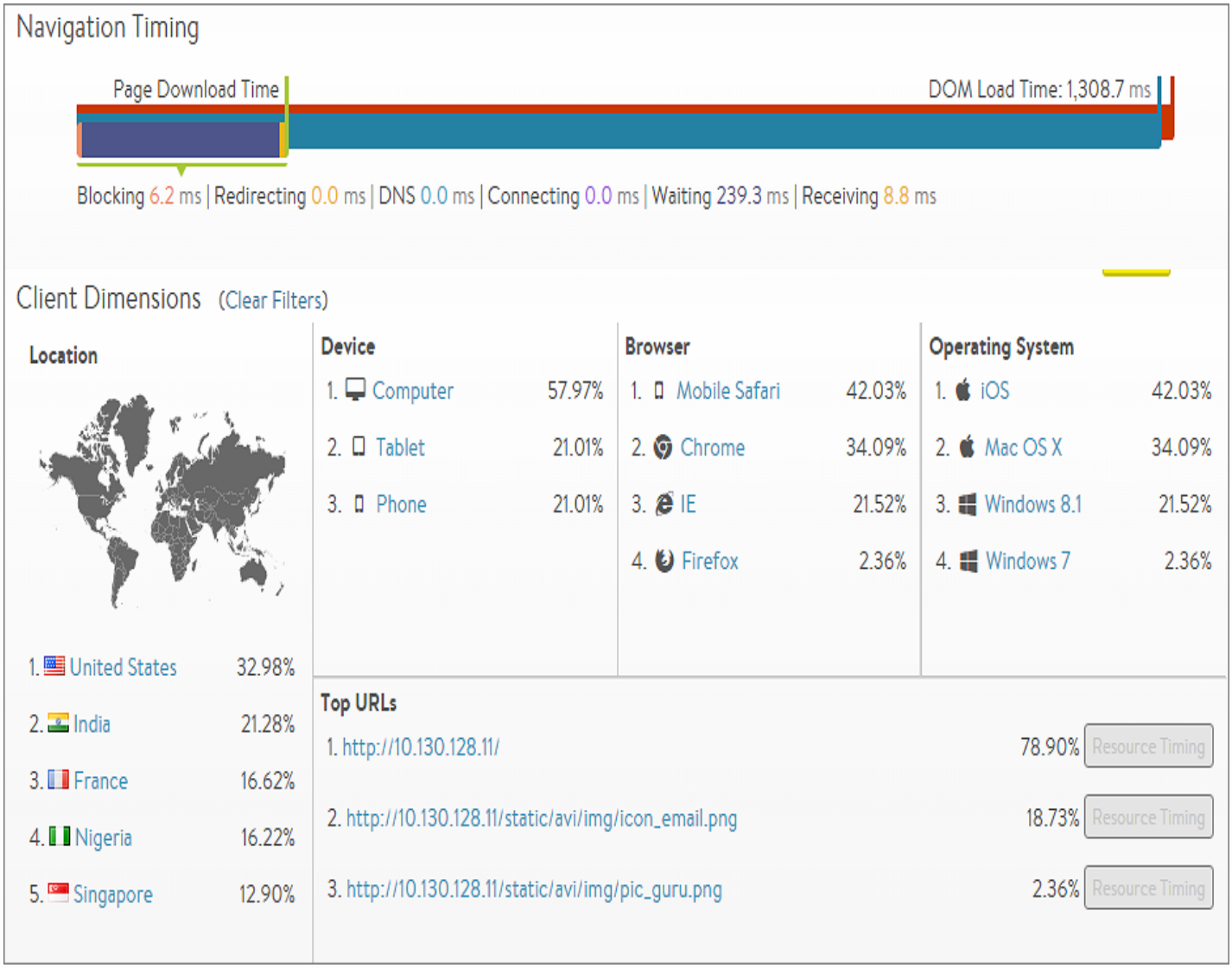
The sidebar tiles show the scores of each of the four subcomponents of the health score, plus the total score. To determine why an object such as a virtual service has a low health score, select one of the first tiles that is showing a subpar score.

This may bring up additional sub-components for the top level metric, such as Pools, connection Apdex, Response Apdex, or others. Select the tile that is showing the worst score. Some tiles may have additional information shown in the main chart section.

The chart pane of the tab shows a timeline graph of the selected metric. Hover the mouse over the chart to display the most significant issue causing degradation of the health score. (This is displayed at the top left of the chart.)

Virtual Service Client Information

The Clients tab for an HTTP virtual service displays information about clients accessing that service. The information shown within the tab is dependent on the setting for client insights. This setting can be modified in the Analytics tab of the Create Virtual Service popup.



Many of the entries on this tab will present expanded data when clicked. For example, clicking Computer in the Devices section filters the Clients tab to display data for clients connecting to the virtual service from a desktop computer (versus phones or tablets). While an entry is highlighted, other sections will have two numbers.

Continuing with the example of highlighting Computers:

- Green or red percentage number shows how much higher or lower the numbers are when filtered for computers compared to without the filter.
- Specific percentage for computers.
- Clicking the selected item again, such as Computer, deselects that item and removes the filter.

The Clients tab presents the following information:

- **End-to-End Timing:** Complete view of client-to-server transaction times. This includes the time required for a client to download an individual HTTP-requested object or to complete a TCP transaction.
- **Page Load Time:** If you have enabled active client insights on the Analytics tab for the virtual service, then Avi Vantage will insert Javascript into a sampling of HTTP responses sent to the client to gather additional information about the client's experience loading the page. This data appears in two bars:
    - **DOM Load Time:** Amount of time from when clients first attempted to load a page served from the virtual service until the client browsers were able to begin rendering the HTML page, even if the entire page was not fully downloaded yet.
    - **Page Load Time:** Total time required to download and render the entire page. These metrics differ from the end-to-end timing in that they are a client's view of the total time taken versus Vantage's view. Thus, they incorporate different metrics. End-to-end timing is the average time to download a single object, whereas page load time is the time to download all objects for the page and then render them in the client's browser. Page load time measures:
        - All HTTP objects, not just a single file
        - HTML errors and timeouts
        - HTTP redirects
        - DNS resolution
        - Blocking, where the browser is waiting for various reasons
        - Rendering time, or the time the browser took to "paint" the page
- **Location:** Client country of origin. Clients connecting from internal IP addresses will display as Internal; however, administrators may create custom locations using IP groups. For example, an IP group can be configured in which "10.1.0.0/16 = West_Campus" and "10.2.0.0/16 = East_Campus". Any clients connecting from these IP ranges will be shown as originating from the custom locations, while all other 10.x.x.x IP addresses will still be classified as Internal. If an IP address exists in multiple places, such as Japan and Tokyo Campus, the more specific address match will be used.

- **Device:** Client device type such as Computer, Tablet, Phone or Other. Vantage derives this information from the client's User-Agent header. Clients that do not present this header or

that present an unknown User-Agent string are categorized as Other.

- **Browser:** Client browser. Vantage tracks Firefox, Safari, Internet Explorer, Opera, and Chrome. This information is sourced from the client's User-Agent string.
- **Operating System:** Client operating system. Vantage tracks Apple MacOS, Apple iOS, Google Android, Microsoft Windows, and Linux. This information is sourced from the client's User-Agent string.
- **Top URLs:** Vantage records the most commonly requested URLs for the virtual service. Clicking the Resource Timing button for one of the URLs opens a detailed waterfall view for that URL and displays a breakdown of the objects referenced by the following:
    - **Selected page**
    - **Ordering** of the objects
    - **Load times** for each object, including DOM load time and total page load time. This information can be very useful for website developers to diagnose slow page load issues.
    - **URL:** Host, path, and query for the requested object.
    - **Initiator:** This object could have been requested because:
        - It was referenced by a link or script.
        - It is an image.
        - It is blank for the root-level page that was clicked by the client.
    - **Start Time:** The time when this object was requested relative to its peers. The first object will always be time 0 and will reference sub-objects, which are requested later. The indicated times are averaged for the recorded requests to this page.
    - **Duration:** Total client download duration. This is a superset of the data points included in the end-to-end timing, because it includes additional metrics beyond the immediate client-to-web-server interaction.
    - **Resource Timing:** Breakdown of the time required to download the page object. The length of the bar indicates the time, and the colors denote the various stages where time was spent. The bars furthest to the left were processed first. Bars that start further to the right are objects that began downloading and processing later.

Resource Timing Metrics

The Resource Timing** breakdown provides the following timing information:

- **DNS:** Time the client spent waiting for the object hostname to resolve.
    - Many objects have the same underlying fully qualified domain name (FQDN) and thus not every object will contain a DNS time.
    - If the client must wait for DNS resolution, then the time could be notably longer.
    - If the client or the client's local DNS server has the hostname cached, DNS time may be imperceptibly small.
- **Blocking:** Time spent while the browser was blocked from further processing. This typically implies the browser was busy reading and processing a page or a script. Failed scripts may cause a head-of-line block while the browser waits to time out the failed process before continuing the page load process.
- **Redirecting:** Time spent following redirects, such as HTTP-to-HTTPS, to a different hostname or path.
- **Connecting:** Time spent connecting with the endpoint (virtual service). This roughly corresponds to the end-to-end timing's client RTT metric.
- **Waiting:** Time spent by the client waiting for the first packet of the response to its request. From the client's perspective, this is a single metric for all objects requested for a page. The End to End Timing graph on the Health tab for the virtual service breaks this into the following separate metrics, recorded on a per object basis:
    - **Server RTT:** The TCP latency between Vantage and the server.
    - **Client RTT:** The TCP latency between Vantage and the client.
    - **Application Response:** The time during which the application server was processing or generating the content.
- **Receiving:** Time spent waiting after the first packet of the response was received until the last packet of the response was received. This is a combination of object size, TCP latency and packet loss, as well as how many round trip times (TCP ACKs) are required. This corresponds to the end-to-end timing chart's data transfer time.
- **Rendering:** The time the browser spent rendering or 'painting' the page.

Virtual Service Security

Avi Vantage continually assesses the health of each virtual service. This health information is available for viewing in both summary and detailed form.

Each virtual service has a health score, which shows the virtual service health as both a color code and a set of numeric scores. The final health score is comprised of a positive performance score and three penalties.

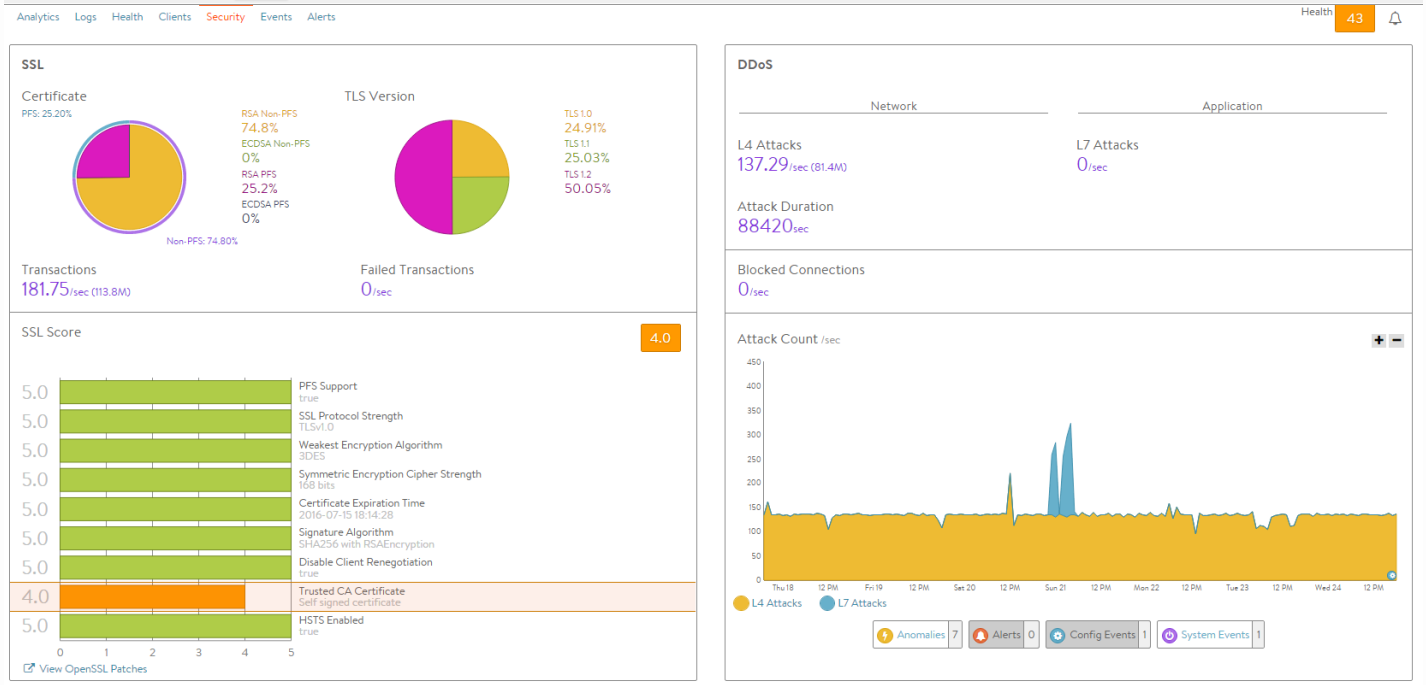
The security penalty provides insight into a current security related issue (such as a current DoS attack) or potential risk (such as SSL configuration which leaves the site vulnerable to the POODLE attack).

Ideally, the security penalty should be zero, which means it is not detracting from the health or risk of a virtual service. A non-zero security penalty may be due to an issue with SSL or a DDoS attack event. This article explores the components that could generate a security penalty.

View Security Insights

To view a security insights for a virtual service:

1. Click the icon of the virtual service shown on the Dashboard.
2. Click the virtual service name.
3. Click Security on the menu bar.



Detailed security information for the virtual service is shown in two panes:

- SSL information (left pane)
- DDoS information (right pane)

Security Insights for the virtual service are organized into the following main categories:

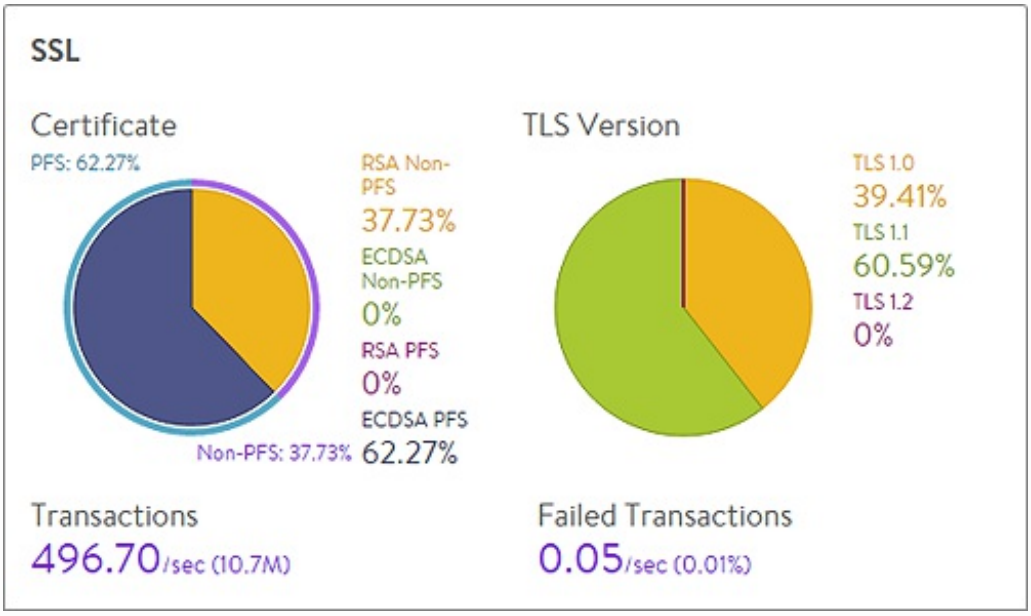
- SSL Distribution
- SSL Score
- DDoS

The following sections describe the data fields listed for each category.

SSL Distribution Insights

The SSL section on the top left of the default security page shows in detail the most relevant SSL data about client connections terminated on Avi Service Engines (SEs) within the selected period of time. If SSL termination is not being performed on the virtual service, this section will have no data.

- **Certificate:** Breaks down the certificate types used by clients during the negotiation phase of SSL session setup. A virtual service can be configured to accept both RSA and EC certificates. Avi Vantage will negotiate whichever type the client supports, with EC as the preference for clients that support both. Depending on the cipher negotiated by the client, RSA and EC may be negotiated with or without Perfect Forward Secrecy.



- **TLS Version:** Breaks down the TLS versions negotiated by clients. Vantage supports TLS, but not the older and less secure SSLv2 or SSLv3.

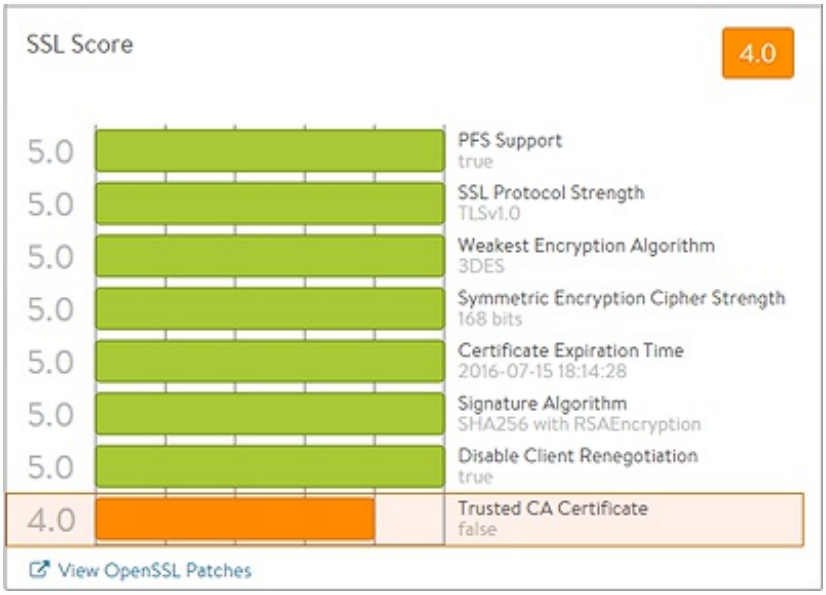
- **Transactions:** The average transactions per second (TPS) for new connections negotiated within the indicated time period. This metric includes both new and reused transactions. This metric is further broken down via the Transactions metric tile from the SSL section of the sidebar tiles.

- **Failed Transactions:** Number of unsuccessful transactions. Typically, transactions may fail either due to clients terminating the negotiation midstream, or because the client and Vantage could not agree on a mutually supported cipher or TLS version. To view individual failed transactions, access the Logs page for the virtual service.

SSL Insights

The SSL Score section on the bottom left of the default security page shows in detail the major factors affecting the SSL Score.

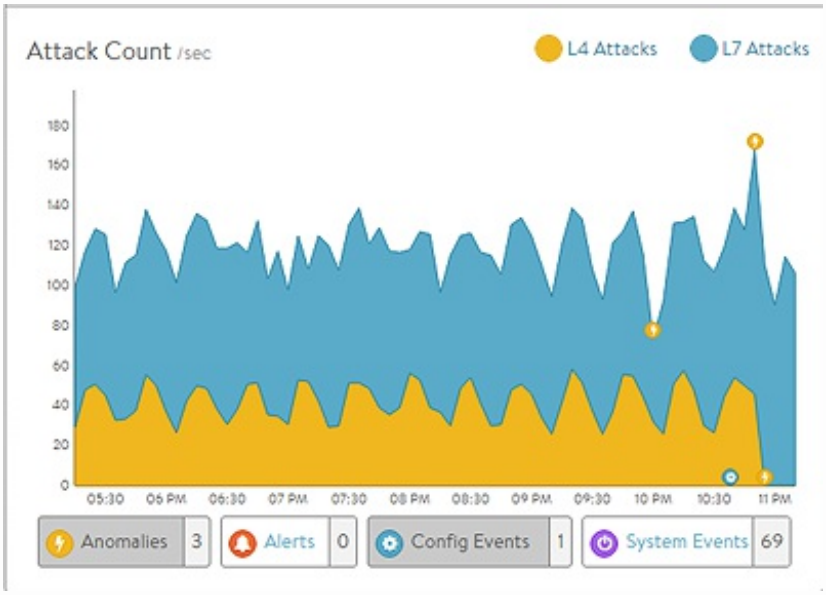
- **PFS Support:** Negatively impacts the virtual service's security score by reducing it if PFS capable ciphers are not enabled in the SSL profile for the virtual service.
- **SSL Performance Strength:** Reduces the score if an insecure SSL/TLS version is enabled.
- **Weakest Encryption Algorithm:** Reduces the score if a weak encryption algorithm is enabled in the SSL profile. See the SSL Profile's security score for more on this.



- **Symmetric Encryption Cipher Strength:** Reduces the score if the cipher suite uses an encryption algorithm with considered insecure.
- **Certificate Expiration Time:** Reduces the score if the certificate is about to expire or has already expired.
- **Signature Algorithm:** Reduces the score if weak hashing algorithms (such as md5 or SHA1) are enabled in the virtual service's SSL profile.
- **Disable Client Renegotiation:** As a best practice, Vantage turns off client SSL renegotiation. This field is non-configurable and therefore does not impact the security score.
- **Trusted CA Certificate:** Reduces the score if the virtual service is using a self-signed certificate.

DDoS Insights

The DDoS section on the right of the default security page breaks down distributed denial of service data for the virtual service into the most relevant layer 4 and layer 7 attack data. See additional documentation on network [attack mitigation][4].



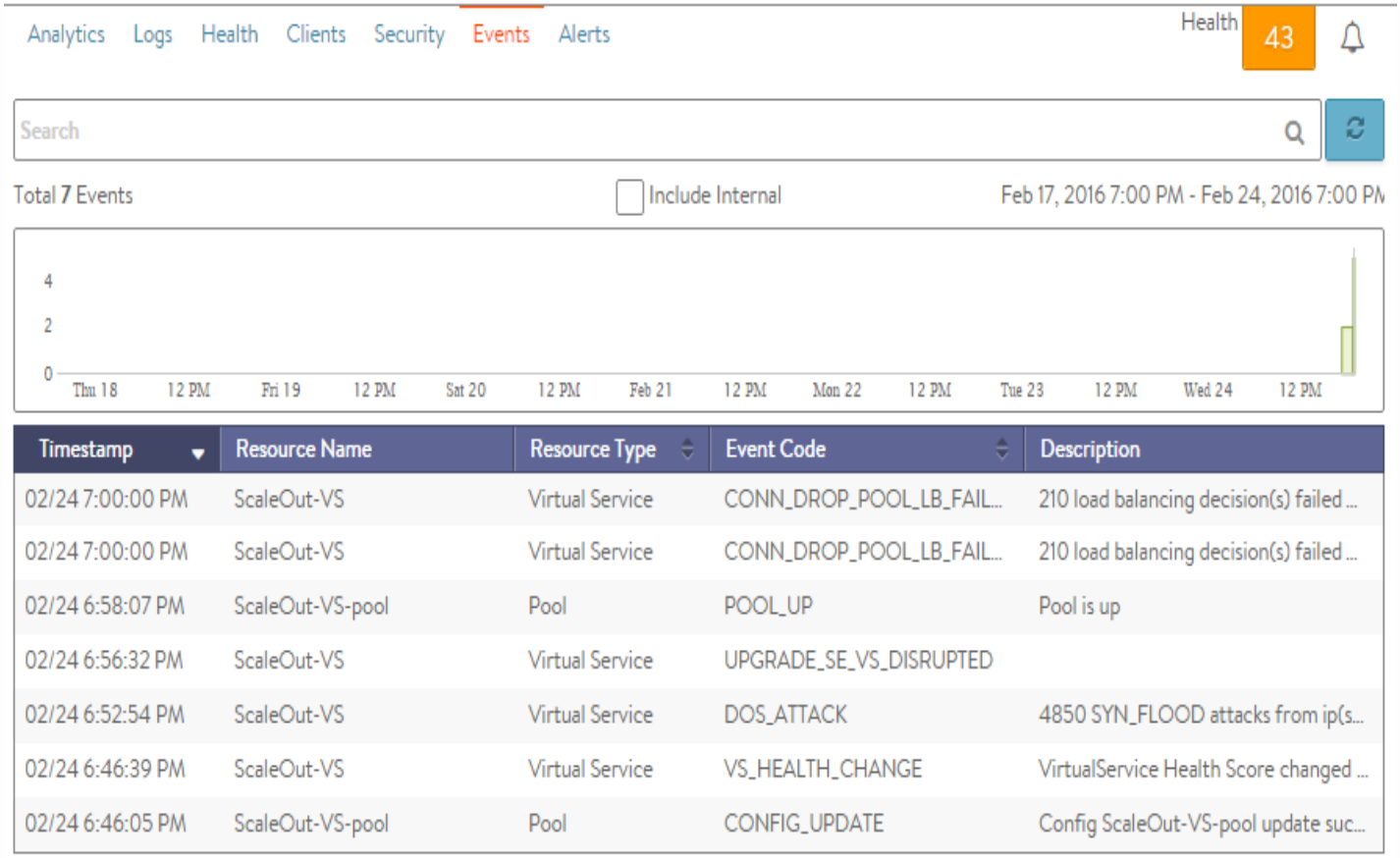
- **L4 Attacks:** The number of network attacks per second, such as IP fragmentation attacks or TCP SYN flood. For the example shown here, each unacknowledged SYN is counted as an attack. (This is the classic signature of the TCP SYN flood attack, a large volume of SYN requests that are not followed by the expected ACKs to complete session setup.)
- **L7 Attacks:** The number of application attacks per second, such as HTTP SlowLoris attacks or request floods. For the example shown here, every request that exceeded the configured request throttle limit is counted as an attack. (See the application profile's DDoS tab for configuring custom layer 7 attack limits.)
- **Attack Duration:** The length of time during which an attack occurred.
- **Blocked Connections:** If an attack was blocked, this is the number of connection attempts blocked.
- **Attack Count:** Shows attacks plotted in a graph over time.

Virtual Service Events Log

Events are recordings or logs, and are used for reactive troubleshooting, but may also be used to trigger alerts, which can be used for proactive actions.

Virtual Service Events

The Events tab presents system-generated events over the time period selected for the virtual service. System events are applicable to the context in which they are viewed. For example, when viewing events for a virtual service, only events that are relevant to that virtual service are displayed.



The top of this tab displays the following items:

- **Search:** The Search field allows you to filter the events using whole words contained within the individual events.
- **Refresh:** Clicking Refresh updates the events displayed for the currently-selected time.
- **Number:** The total number of events being displayed. The date/time range of those events appear beneath the Search field on the left.
- **Clear Selected:** If filters have been added to the Search field, clicking the Clear Selected (X) icon on the right side of the search bar will remove those filters. Each active search filter will also contain an X that you can click to remove the specific filter.
- **Histogram:** The histogram shows the number of events over the period of time selected. The X-axis is time, while the Y-axis is the number of events during that bar's period of time.
    - Hovering the cursor over a histogram bar displays the number of entries represented by that bar, or period of time.
    - Click and drag inside the histogram to refine the date/time period which further filters the events shown. When drilling in on the time in the histogram, a Zoom to selected link appears above the histogram. This expands the drilled in time to expand to the width of the histogram, and also changes the Displaying pull-down menu to Custom. To return to

the previously selected time period, use the Displaying pull-down menu.

The table at the bottom of the Events tab displays the events that matched the current time window and any potential filters. The following information appears for each event:

- **Timestamp:** Date and time the event occurred. Highlighting a section of the histogram allows further filtering of events within a smaller time window.
- **Event Type:** This may be one of the following:
  - **System:** System events are generated by Avi Vantage to indicate a potential issue or create an informational record, such as VS_Down.
  - **Configuration:** Configuration events track changes to the Avi Vantage configuration. These changes may be made by an administrator (through the web interface, API, or CLI), or by automated policies.
- **Resource Name:** Name of the object related to the event, such as the pool, virtual service, SE, or Controller.
- **Event Code:** A short event definition, such as Config_Action or Server_Down.
- **Description:** A complete event definition. For configuration events, the description will also show the username that made the change.
- **Expand/Contract:** Clicking the plus (+) or minus sign (-) for an event log either expands or contracts that event log. Clicking the + and – icons in the table header expands and collapses all entries in this tab.

For configuration events, expanding the event displays a difference comparison between the previous and new configurations.

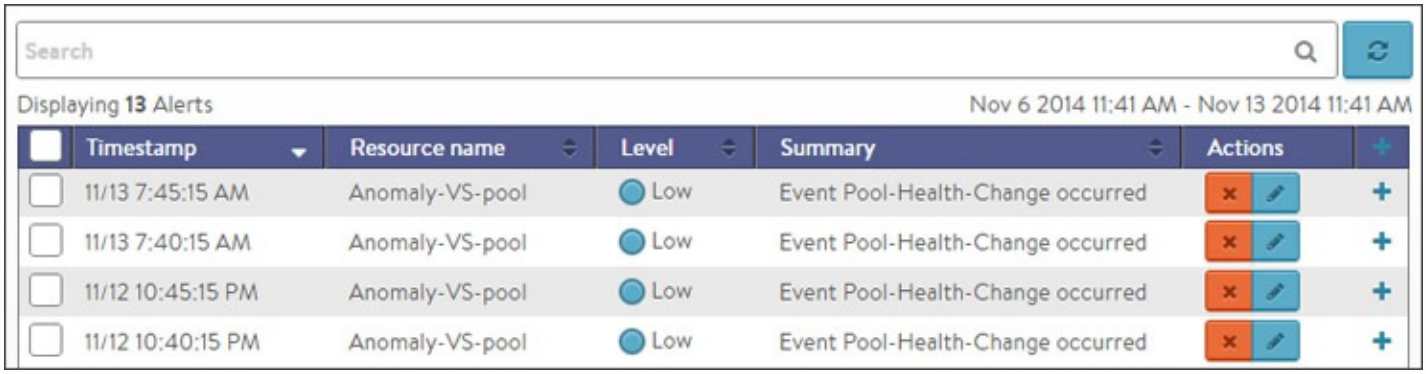
- New fields will appear highlighted in green in the new configuration.
- Removed fields will appear highlighted in red.
- Changed fields will show highlighted in yellow.

Virtual Service Alerts

Alerts act as filters that provide notification for prioritized events or combinations of events. The virtual service alerts tab highlights user-specified events for the virtual service for the selected time period.

Alert Actions

Alert actions may be configured to provide proactive notifications via Syslog, SNMP traps, email, or ControlScripts. This is done via the the Alerts tab of the Operations section of the Vantage web interface, specifically in the Alert Action sub page.



The top of this tab shows the following items:

- **Search:** The Search field allows you to filter the alerts using whole words contained within the individual alerts.
- **Refresh:** Clicking Refresh updates the alerts displayed for the currently-selected time.
- **Number:** The total number of alerts being displayed. The date/time range of those alerts appear beneath the Search field on the left.
- **Dismiss:** Select one or more alerts from the table below then click Dismiss to remove the alert from the list. Alerts are transitory, meaning they will eventually and automatically expire. Their intent is to notify an administrator of an issue, rather than being the definitive record for issues. Alerts are based on events, and the parent event will still be in the events record.

Alert Table

The table at the bottom of the Alerts tab displays the following alert details:

- **Timestamp:** Date and time when the alert was triggered. Changing the time interval using the Displaying pull-down menu may potentially show more alerts.
- **Resource Name:** Name of the object that is the subject of the alert, such as a server or virtual service.
- **Level:** Severity level of the alert, which can be High, Medium, or Low. Specific notifications can be set up for the different levels of alerts via the Administration page’s Alerts Overlay.

- **Summary:** Summarized description of the alert.
- **Action:** Click the appropriate button to act on the alert:
  - **Dismiss:** Clicking the red X dismisses the alert and removes it from the list of displayed alerts.
  - **Edit:** Clicking the blue pencil icon opens the Edit Alert Config popup for the alert configuration that triggered this alert. This can include a verbose and customized description of the alert or allow an administrator to alter settings such as the severity of the alert.
  - **Expand/Contract:** Clicking the plus (+) or minus sign (-) for an event log either expands or contracts that event log to display more information. Clicking the + and – icon in the table header expands and collapses all entries in this tab

Create a Virtual Service

A new virtual service may be created via either the basic or advanced mode. In basic mode, many features are not exposed during the initial setup. After the virtual service has been created via basic mode, the options shown while editing are the same as advanced mode, regardless which mode was initially used. While basic mode may have been used to create the virtual service, it does not preclude access to any advanced features.

- **Basic Mode**: This mode is strongly recommended for the vast majority of normal use cases. It requires minimal user input and relies on pre-defined configurations for the virtual service that should be applicable for most applications. Creating a new virtual service in basic mode can be accomplished within a single popup window.
- **Advanced Mode**: This mode requires additional user input and is recommended when requiring access to less common features, such as policy rules or customized analytics settings. This mode may also be used to configure a virtual service for multiple ports or network protocols.

Virtual Service Creation: Basic Mode

Creating a virtual service in basic mode enables quick creation of all required objects, in particular the pool containing servers. The automatically created pool will be given the name *virtual-service-name*-pool, where *virtual-service-name* is the name of the virtual service being created.

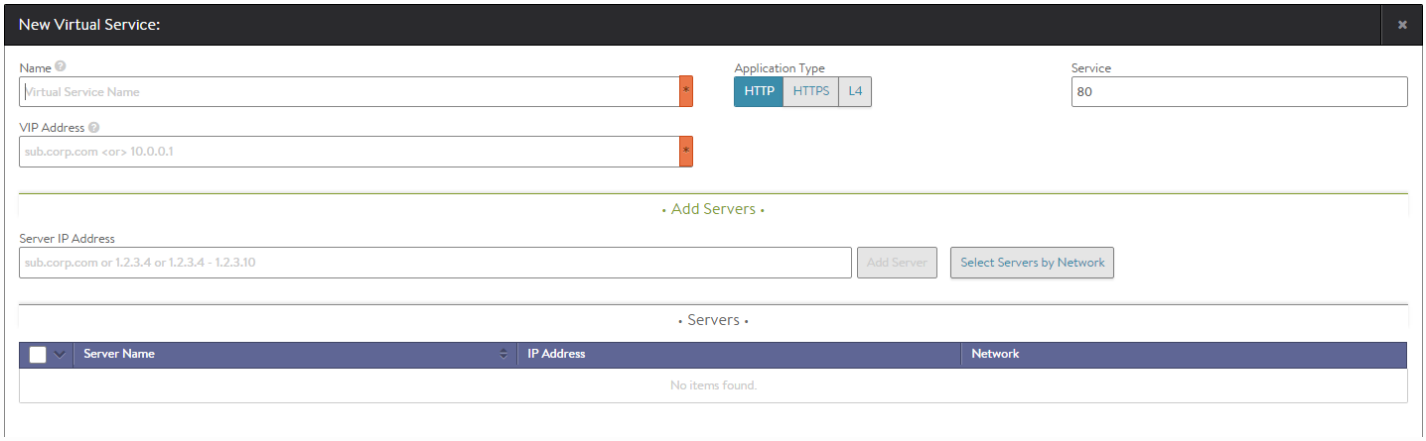
To assign an existing pool to a virtual service, do one of the following:

- Create the virtual service in basic mode with the server field left blank, then edit the virtual service to assign the desired pool.
- Create the virtual service in advanced mode.

There are some limitations to note when creating a virtual service in basic mode:

- The virtual service will be assigned the default application profile, either HTTP, HTTPS, or Layer 4.
- The virtual service will be assigned the default TCP/UDP profile, TCP Proxy.
- After creating the virtual service, Real Time Metrics are enabled for the first 60 minutes and Full Client Logs are enabled for the first 30 minutes. After this time period lapses, the metrics will revert to five minute update intervals and logs will capture errors only. After the virtual service creation process is complete, editing the virtual service exposes all advanced functions, allowing any configuration setting to be altered.

Create a new virtual service via Applications > Virtual Service > New > Basic Mode:



- **Select Cloud:** If Vantage is configured for multiple cloud environments, such as VMware and Amazon Web Services (AWS), the cloud for the virtual service deployment must be selected. If Vantage exists in a single environment, this question is skipped.
- **Name:** Provide a unique name for the new virtual service. (For clouds configured with Cisco ACI as the networking fabric, the name becomes a pull-down menu populated with the list of existing Service Graphs that are not currently assigned to a virtual service.) </p>
- **VIP Address:** Enter either the DNS resolvable name or an IP address for the virtual service. When using a DNS name, Vantage will attempt to resolve the name and populate the IP Address field if successful. If the name resolves to more than one IP address, select the address to use from the list shown. If the name cannot be resolved, it will appear in red. See Virtual Service Using DNS for VIP Address.</p>

- **Application Type:** Select from the common app types. To use a different application profile or type, consider creating the virtual service via advanced mode.</p>
    - **HTTP:** The virtual service will listen for non-secure Layer 7 HTTP. Selecting this option auto-populates the Service port field to 80. Override the default with any valid port number; however, clients will need to include the port number when accessing this virtual service. Browsers default to automatically append the standard port 80 to HTTP requests. Selecting HTTP enables an HTTP application profile for the virtual service. This allows Vantage to proxy HTTP requests and responses for better visibility, security, acceleration, and availability.
    - **HTTPS:** The virtual service will listen for secure HTTPS. Selecting this option auto-populates port 443 in the Service port field. Override this default with any valid service port number; however, clients will need to include the port number when accessing this virtual service as browsers automatically append the standard port 443 to HTTPS requests. When selecting HTTPS, use the Certificate pull-down menu to reference an existing certificate or create a new self-signed certificate. A self-signed certificate will be created with the same name as the virtual service and will be an RSA 2048 bit cert and key. The certificate can be swapped out later if a valid certificate is not yet available at time of virtual service creation.
    - **L4:** The virtual service will listen for Layer 4 requests on the port you specify in the Service port field. Select this option to use the virtual service for non-HTTP applications, such as DNS, mail, or a database.
    - **Service:** Select the service port to listen for connections. Basic mode supports configuring only one port. For multiple service ports or ranges, edit the virtual service after creation.
- **Add Servers:** Add one or more servers to the new virtual service using any combination of the following methods:
    - **Add by Server IP Address**: Enter the IP address for the server to be included in the Address field, and then click the Add Server button. A range of IP addresses may be entered using a dash character, such as 10.0.0.1–10.0.0.10.
    - **Add by DNS Resolvable Server Name**: Enter the name of the server to be included in the Address field. If the server name successfully resolves, the IP address will be shown and the Add Server button will change to green. Click the Add Server button. See Add Servers by DNS.
- **Select Servers by Network**: This option is only available if Vantage has read or write access to the cloud environment. Click the Select Servers by Network button to open a list of

reachable networks. Selecting a network opens a list of the servers (virtual machines) available on that network. Click the green Add Servers button to add the selected servers to the virtual service. See Select Servers by Network.

- **Save**: Click Save to finish. The new virtual service is now displayed in the Applications Dashboard and Virtual Services Page.

The virtual service will be assigned automatically to a Service Engine. If an available SE already exists, the virtual service will be deploy and be ready to accept traffic. If a new SE must be created, it may take a few minutes before it is ready.

The color of the virtual service name icon indicates the status of the new virtual service. Hovering the mouse over the virtual service health score will convey more information regarding the status of the new virtual service.

In some environments, Avi Vantage may require additional networking information such as IP addresses or clarification of desired networks, subnets, or Port Groups to use prior to a new Service Engine creation. The UI will prompt for additional information if this is required.

Virtual Service Creation: Advanced Mode

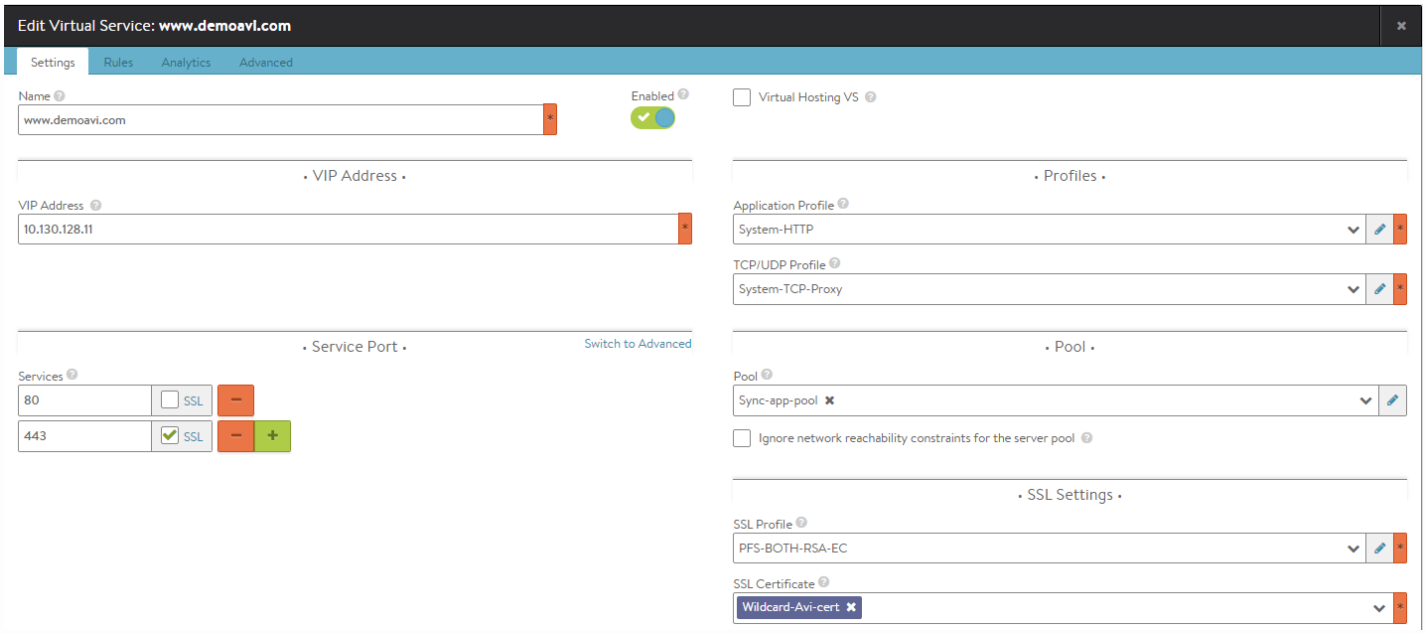
The Create Virtual Service popup and Edit Virtual Service popup share the same advanced mode interface, which has the following tabs:

- Settings Tab
- Rules Tab
- Analytics Tab
- Advanced Tab

When creating a new virtual service in Advanced mode, you will be guided through the tabs in order, starting with the Settings tab and concluding with the Advanced tab. If you are editing an existing virtual service, you will be able to jump to any tab non-linearly.

Advanced Virtual Service Creation: Settings Tab

The Settings tab of the Create/Edit Virtual Service popup covers the basic network settings for the virtual service.



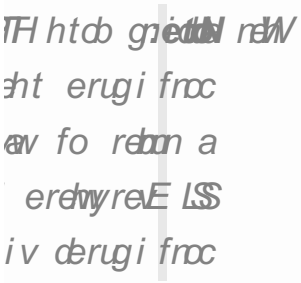
To add or edit virtual server settings:

- **Select Cloud:** If Vantage is configured for multiple cloud environments, such as VMware and Amazon Web Services (AWS), then the desired cloud for the virtual service must be selected during virtual service creation. If Vantage exists in a single environment, this question is skipped.
- **Name:** Provide a unique name for the new virtual service. For clouds configured for Cisco ACI for the networking fabric, the name becomes a pulldown menu populated with the list of existing Service Graphs that are not assigned to a virtual service.
- **Enabled:** The toggle icon enables (green) and disables (red) the virtual service.
    - When enabled (green icon), the virtual service will accept and process traffic normally.
    - When disabled (red icon), the virtual service will not accept any new connections. Existing concurrent connections will be terminated, and the virtual service will be unassociated from all Service Engines. No health monitoring is performed for disabled virtual services. See Disable a Virtual Service.
    - **VIP Address:** Enter either the DNS resolvable name or an IP address for the virtual service. When using a domain name, Vantage will attempt to resolve that name to populate the IP Address field. If the name resolves to more than one IP address, select the address to use from the list shown. If the name cannot be resolved, it will appear in red. See Virtual Service Using DNS for VIP Address.
    - **AWS**
        - **Elastic IP Address For External Client Access**: Will allocate a public IP address

that AWS will use to NAT to the VIP address. This may be automatically allocated by AWS, or a static address may be specified.

- **Network for VIP Address Allocation**: Specify the network for client reachability. An elastic IP address will be automatically chosen within the selected network.
- **OpenStack**
- **Floating IP Address For External Client Access**: Optionally specify an external IP address that will be exposed to clients outside of OpenStack. This may be automatically allocated by OpenStack, or a static address may be specified.
- **VIP Address**: When set to auto-allocate, OpenStack will provide the VIP address once a VIP network has been defined.

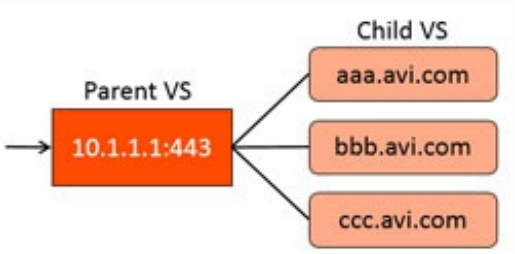
- **Services:** Define the listening Service port of the virtual service. This may be configured via the default basic mode, or via switching to advanced mode.
  - **Services Basic:** Enter a single service port number, such as 80 for HTTP. Multiple service ports may be added by clicking the green plus icon.



  - **Service Advanced:** In advanced mode, services may be added individually, or as a range, such as 5000 to 10000. This can be useful for older apps such as CORBA. The virtual service default TCP/UDP profile may be overridden on a per-service port basis. This can be beneficial for protocols such as DNS, RADIUS, or Syslog that may use both TCP and UDP at the same time. To change the protocol, select Override TCP/UDP and select a profile for this service port.
- **SSL:** If the virtual service will be decrypting client SSL traffic on this port, check the SSL checkbox. This brings up two additional fields:
  - **SSL Profile:** Collection of settings that determine the preferred ciphers and versions of SSL that Vantage will use when negotiating with clients.
  - **SSL Certificate:** Either select an existing SSL certificate or create a new certificate. Vantage supports both RSA and Elliptic Curve (EC) certificates, and both may be configured for the same virtual service. To use both RSA and EC, simply add a certificate

for each type in this field. Vantage will prefer the EC certificate because EC is newer, more secure, and faster. Most any browser from the past decade will support EC.

- **Virtual Hosting VS:** When selected, this virtual service participates in virtual hosting via SSL's Server Name Indication (SNI). This allows a single SSL decrypting virtual service IP:Port to forward traffic to different internal virtual services based on the name of the site requested by the client. The virtual hosting VS must be either a parent or a child.
    - **Parent:** The parent virtual service is external facing, and owns the listener IP address, service port, network profile, and SSL profile. Specifying a pool for the parent is optional, and will only be used if no child virtual service matches a client request. The SSL certificate may be a wildcard certificate or a specific domain name. The parent's SSL certificate will only be used if the client's request does not match a child virtual service domain. The parent virtual service will receive all new client TCP connections, which will be reflected in the statistics. The connection is internally handed off to a child virtual service, so subsequent metrics such as concurrent connections, throughput, requests, logs and other stats will only be shown on the child virtual service.

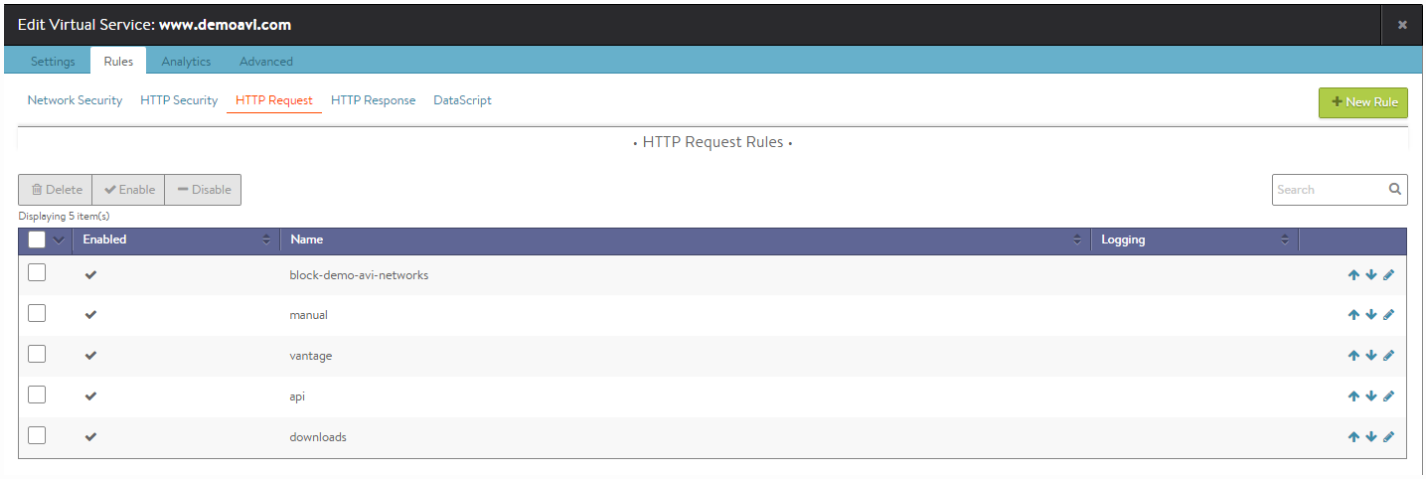


        - **Child:** The child virtual service does not have an IP address or service port. Instead, it points to a parent virtual service, which must be created first. The domain name is a fully qualified name requested by the SNI-enabled client within the SSL handshake. The parent matches the client request with the child's domain name. It does not match against the configured SSL certificate. If no child matches the client request, the parent's SSL certificate and pool are used.
- **Application Profile:** Use the application profile pull-down menu to select how the virtual service will listen for client requests. Selecting an HTTP profile will allow the virtual service to proxy HTTP and provide functionality specific to that protocol. An HTTP-based application profile will also require the TCP/UDP profile (described further below) to be set to TCP Proxy. For non-HTTP applications, select an L4 application profile.
- **TCP/UDP Profile:** Use the TCP/UDP profile to set the virtual service to listen for UDP Fast Path, TCP Fast Path or TCP Proxy. If the application profile is set to an HTTP profile, the TCP/UDP profile must be set to proxy the TCP connections.
- **Pool:** Select the pool to use for this virtual service using the Pool pull-down menu. A pool may only be referenced by one virtual service. Pools that are already in use will not appear on this list. Click the green Create button to create a new pool. If the virtual service requires sending traffic to multiple pools, configure content switching with an HTTP request policy.

- **Ignore Reachability:** If the pool contains servers in networks unknown or inaccessible to Vantage, the Controller is unable to place the new virtual service on a SE, as it doesn't know which SE has the best reachability. This requires manually choosing the virtual service placement by the Vantage user. Selecting *Ignore Pool servers network reachability constraints for Virtual Service placement* will allow the Controller to place the virtual service, even though some or all servers in the pool may be inaccessible. For example, a Vantage user may select this option while creating the virtual service, and later configure a static route to access the servers.

Advanced Virtual Service Creation: Rules Tab

The rules tab allows you to define policies or DataScripts for the virtual service.



DataScripts and policies consists of one or more rules that control the flow of connections or requests through the virtual service to control security, client request attributes, or server response attributes.

Each rule is a match/action pair that uses if/then logic: If something is true, then it matches the rule and corresponding actions will be performed.

Policies are simple GUI-based, wizard driven logic, whereas DataScript allows more powerful manipulation using Vantage's Lua-based scripting language.

For an in-depth explanation of policy configuration, see the Virtual Service Policies section. For help writing DataScripts, see the Avi DataScript Guide.

Policies are specific to the virtual service they are configured within. They consist of a simple "if match is true, then perform action" logic. Each line of match/action logic is a rule.

A policy can contain multiple rules, in which case the rules are processed in the order they are listed in the policy's configuration in the web interface.

- A rule may have just an action without a match, which means that every connection or

request will be an automatic match.

- A rule may contain many matches and/or many actions.

Policy Types

Vantage supports the following types of policies.

- Network security policy: explicitly allows or blocks traffic based on network (TCP/UDP) information.
- HTTP security policy: performs defined actions such as allow/deny, redirect to HTTPS, or respond with a static page.
- HTTP request policy: manipulates requests or performs content switching by allowing customized actions based on client HTTP requests.
- HTTP response policy: evaluates responses from the server and can be used to modify the server's response headers. This type of policy is most often used in conjunction with HTTP request policies to provide web site name modification similar to Apache's Mod_Proxypass.

Policy Processing Order

Policies and their rules will be processed in the order listed above.

Adding a Policy

To add a new policy:

1. **Policy Type:** Select the policy type to add by selecting a category.
2. **Create** a new rule by clicking the green Add Rule button and entering the following information for the new rule:
    - **Enable/Disable:** By default, the new rule will be enabled (on). Clicking the on/off switch toggles this rule On (green icon) or Off (gray icon). Disabled rules have no effect on traffic.
    - **Name:** Either enter a unique name for the rule in the Rule Name field or leave the default system generated name in place.
    - **Logging:** Check the Logging checkbox to enable logging for this rule. When enabled, Vantage will generate a log for any connection or request that matches the rule's match criteria. If a virtual service is already set to log all connections or requests, Avi Vantage will not create a duplicate log. The logs will be flagged with a new field for the policy type

and rule name that matched. When viewing the policy logs within the Logs tab of the Virtual Service Details page the logs will be included as part of the Non-Significant (Full Client Logs) option unless the specific connection or request is an error and qualifies as Significant.

- **Match:** Add one or more match criteria using the matching rules pull-down menu. The match options will vary depending on the context defined by the policy type being creating. If a rule is not given a match, then all connections or requests will be considered matched. For complete descriptions of the match options available for each policy type, see the help for that type.
- **Action:** Add one or more Actions to be taken when the rule is triggered using the Action pull-down menu. The available options will vary depending on the type of rule you are creating.
- **Save:** Click the Save Rule button at the bottom of the popup to save the new rule.

3. **Order:** Rules within a policy are enforced in the order in which they appear in the list. For example, if you add a rule to close a connection based on a client's IP address followed by a rule that redirects HTTP to HTTPS, Vantage will close the connection without redirecting the request. To alter, the order in which rules are applied, click the up and down arrow icons until the rules are in the preferred order.
4. **Next/Save:**
   - If you are creating a new virtual service, then click Next to proceed to the Analytics tab.
   - If you are editing an existing virtual service, either select the next tab to continue edits or click Save to save your changes and exit back to the Virtual Services page.

DataScript

DataScripts are global objects that may be applied to multiple virtual services. More than one DataScript may be applied to a single virtual service. The order of the DataScripts is important. For instance, if the first script enforces a blacklist that may drop clients, and the second script redirects clients to another site, a client that is dropped by the first rule will never be evaluated against the second. Use the up and down arrows next to the DataScript to reorder. See also Execution Priority.

To add a DataScript to the virtual service: from the DataScript tab, select new DataScript to add a rule to the virtual service. Select a rule from the Script to Execute menu. Choose an existing script or create a new one. Select Save Script to add this DataScript to the virtual service.

The scripts are maintained in the Templates > Scripts > DataScripts section of the web interface.

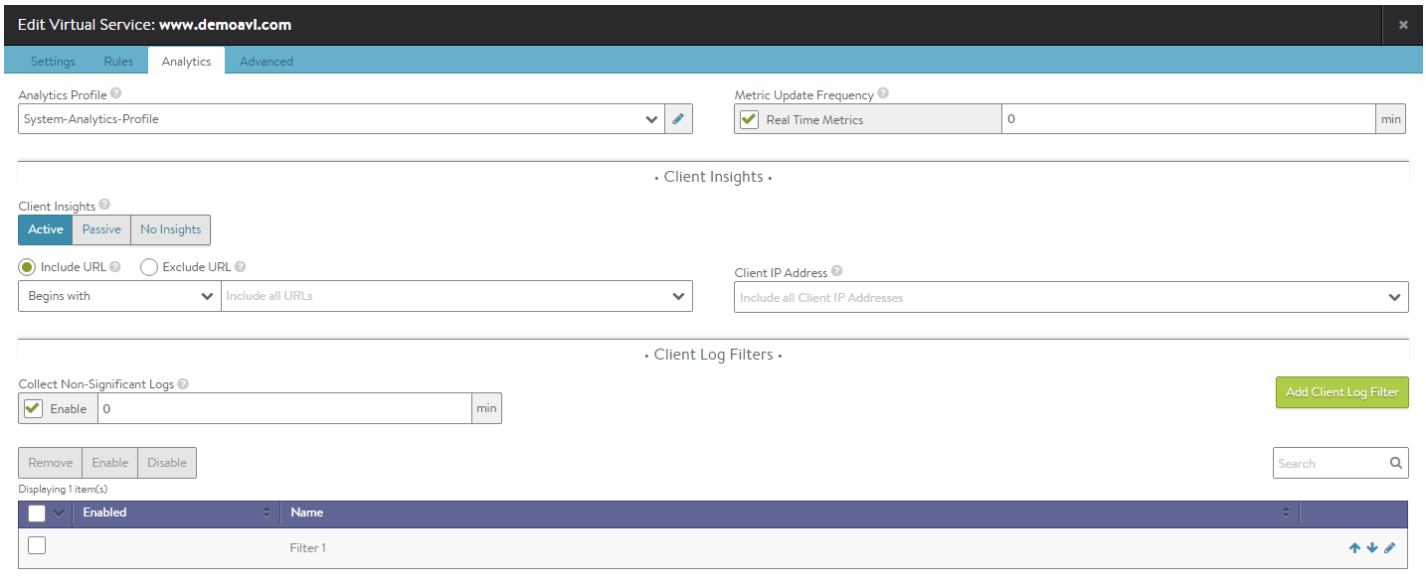
To create a DataScript, start with the overview of DataScripts, which walks through the high level objects and has links to numerous sample scripts.

Advanced Virtual Service Creation: Analytics Tab

The Analytics tab of the advanced virtual service configuration wizard defines how analytics data is captured. These settings control:

- Thresholds for defining client experience and the resulting impact on the end-to-end timing and health score
- Level of metrics collection
- Logging behavior

Analytics



- **Analytics Profile:** Select an analytics profile from the pull-down menu. This profile determines the thresholds for determining client experience. It also defines errors that can be tailored to ignore certain behavior that may not be an error for a site, such as an HTTP 401 (authentication required) response code. Avi Vantage uses errors and client experience thresholds to determine the health score of the virtual service and may generate Significant Log entries for any issues that arise.
- **Real Time Metrics:** The Analytics tab of the Virtual Service Details page presents a number of metrics, such as Throughput and Requests. Avi Vantage updates these metrics periodically, either the default of 5 minutes, or via the Real Time Metrics option, every 15 seconds. Due to the increased resources to maintain more detailed metrics, Real Time Metrics may be configured to gather metrics more aggressively for a limited period of time. A time of 0 enables the aggressive metrics collection indefinitely. Setting this to a shorter time

value, such as 30 minutes, means Real Time Metrics collection will revert back to the slower polling once the specified time has elapsed, for instance, after you have completed any troubleshooting session.

Client Insights

Data about connecting clients may be captured. Specific clients may be included or excluded via the Include URL, client IP address, and exclude URL options. This data is used in pages such as the Virtual Service > Clients page.

Select the level client insights to collect:

- **No Insights:** No client insights are recorded for this virtual service.
- **Passive:** Record data passively flowing through the virtual service. This option enables recording of the End-to-End Timing and client's location. For HTTP virtual services, device, browser, operating system, and top URLs metrics are also included. No agents or changes are made to client or server traffic.
- **Active:** For HTTP, active mode goes further by inserting an industry standard JavaScript query into a small number of server responses to provide HTTP navigation and resource timing. Client browsers transparently return additional information about their experience loading the web page. Avi Vantage uses this information to populate the Navigation Timing and Top URL Timing metrics on the Virtual Service > Clients page. A maximum of one HTTP web page per second will be selected for generating the sampled data.

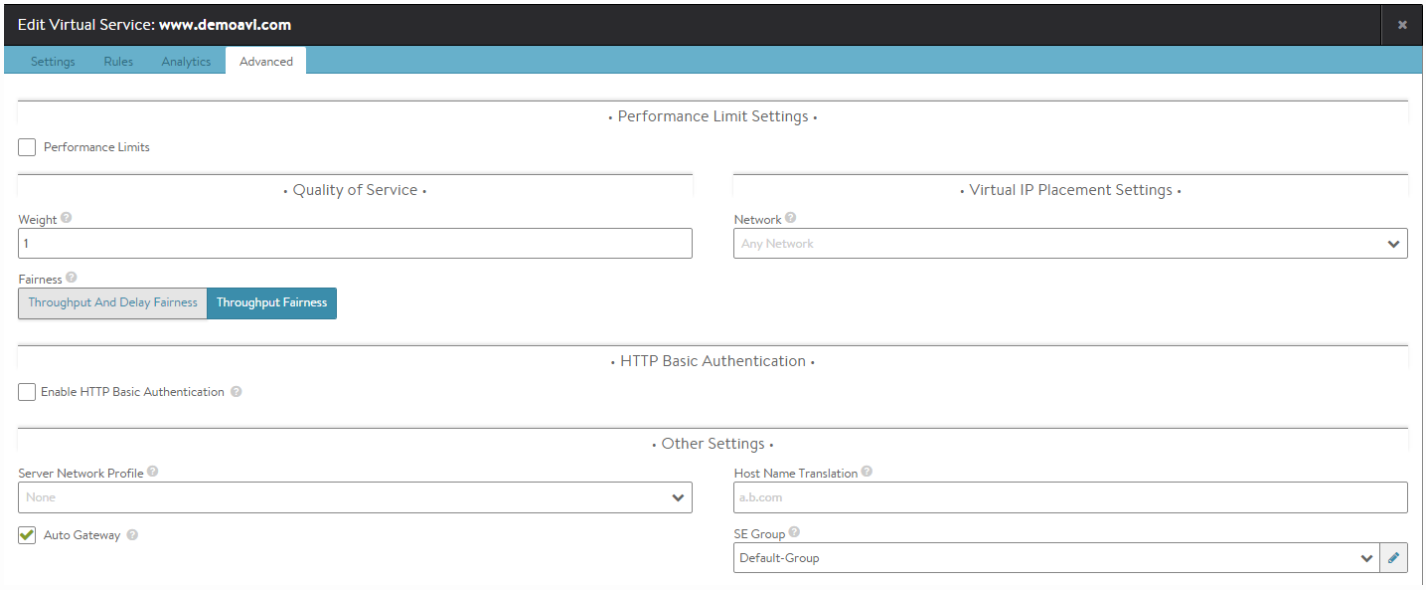
Non-significant Logs

By default, Vantage captures and logs any errors, such as reset connections or HTTP 500 errors, which are classified as Significant logs. To capture full connection logs or HTTP client request logs, enable the Full Client Logs option. View logs in the Logs tab of the Virtual Service Details page. From the logs page, you may choose to view logs filtered by Significant logs (such as errors) or Non-Significant Logs, which includes logs generated via the Full Client Logs option as well as any logs generated via a policy or DataScript.

- **Enable:** Because Non-Significant Logs consume additional resources on busy Vantage systems, by default Non-Significant Logs is disabled. To enable indefinitely, select the Enable checkbox and set the time to 0. Setting the time to a number greater than 0, such as 30, will provide the additional logs for 30 minutes before reverting back to Significant Logs collection only.
- **Client Log Filter:** Rather than log every request or connection, log collection can be filtered to relevant IP addresses and HTTP URLs for a specified duration of time.
- **All Headers:** This option, nested within a client log filter, will log all client request and server response headers, rather than just the common headers which are logged normally. This will increase the size of each log, which reduces the total number Vantage can collect, index, and store.
- **Matching Filter:** If both an IP Address and URL are provided, they must both match for the filter to create the log. IP addresses can be specific addresses such as 10.1.1.1, networks such as 10.0.0.0/8, or IP groups. Setting the order of multiple filters has no effect on the logging as all must match.

Advanced Virtual Service Creation: Advanced Tab

The Advanced tab specifies advanced and optional settings for the virtual service.



To specify performance limits for a virtual service, expand the list of options via the Enable option. The limits applied are for this virtual service only, and are based on an aggregate of all clients. Limit options on a per client basis may be set via the application profile's DDoS tab, which would be inherited by any virtual service using the profile. Additional per client limits may be configured via policies or DataScripts.

Performance Limits

See Rate Shaping and Throttling for additional options.

- **Rate Limit Number of New TCP Connections:**
  - **Threshold:** Set the maximum threshold of new connections from all clients that may be created for this virtual service over the configured time period.
  - **Time Period:** The time, in seconds, within which the threshold must not be exceeded. Set to 0 for infinite.
  - **Action:**
    - Report only: generate logs for excessive connections, but otherwise allow them to proceed as normal.
    - Drop SYN Packets: silently discard the client SYN.
    - Send TCP RST: send a connection reset back to the client.
- **Rate Limit Number of New HTTP Requests:**
  - **Threshold:** Set the maximum threshold of HTTP requests from all clients for this virtual service over the configured time period.
  - **Time Period:** The time, in seconds, within which the threshold must not be exceeded. Set to 0 for infinite.
  - **Action:**
    - Report only: generate logs for excessive connections, but otherwise allow them to proceed as normal.
    - Close TCP Connection: send a connection RST to the client.
    - Send Local Response: allow Vantage to respond with a simple web page and appropriate status code.
    - Send HTTP Redirect: forward (redirect) excessive requests to another URI or destination.
- **Max Throughput:** Specify the maximum amount of bandwidth in Mbps Vantage will allow through the virtual service. Traffic that exceeds this limit will be discarded and may require retransmission by either the client or server. When this limit is set, the Analytics tab of the Virtual Service Details page will show a dashed horizontal line on the throughput metric chart that shows the current throughput versus the maximum allowed throughput.

- **Max Concurrent Connections:** Specify the maximum number of concurrent open connections. Connection attempts that exceed this number will be reset (TCP) or dropped (UDP) until the total number of concurrent connections falls below the threshold. When this limit is set, the Analytics tab for the virtual service will show a dashed horizontal line on the open connections metric that shows the current open connections versus the maximum number of allowed connections.

Quality of Service

- **Weight**: When multiple virtual services are sharing the same Service Engine, and one or more virtual services are consuming most or all available bandwidth, the remaining virtual services may become starved for bandwidth. Bandwidth is loosely defined as packets per second through the SE's hypervisor, saturation of the physical interface of the host server, or similar network constrictions. Avi Vantage provides throughput fairness to ensure that all virtual services are guaranteed some amount of bandwidth. The first option is the weight, which may be set to 1, 2, 4, or 8. This settings determines the relative ratio of traffic that each virtual service should be able to send relative to its peers. This setting is only applicable if there is network congestion, and only for packets sent from the Service Engine. See also throughput fairness field below.

- **Fairness**: Fairness determines the algorithm used to ensure each virtual service is able to send traffic when the Service Engine is experiencing network congestion. The throughput fairness algorithm takes the virtual service weight (see above) into account to achieve this. Throughput and delay fairness is a more thorough algorithm to accomplish the same task. It consumes greater CPU on the Service Engine when there are larger numbers of virtual services. This option is only recommended for latency sensitive protocols.

Placement Settings

- **Network**: Avi Vantage will automatically place the virtual service into a VMware port group for a virtualized environment based on its IP address; however, some environments may have a subnet which spans multiple port groups. The network setting allows the virtual service IP address to be manually set to a specific port group, rather than one automatically picked by Avi Vantage.

HTTP Basic Authentication

- **Basic Realm**: When a request to authenticate is presented to a client, the basic realm indicates to the client which realm they are accessing.

- **Include URL**: If the client request is matches the URI path specified here, then force the user

to authenticate if they have not done so yet.

- **Auth Profile**: Authentication settings such LDAP servers are maintained in the reusable *Templates > Auth Profile*.

Other Settings

- **Server Network Profile**: By default, a new virtual service will use the same TCP/UDP profile specified in the Settings tab for both the client and the server side of a TCP proxied connection. You may override this setting for the connection between the Service Engine and the server by specifying a different TCP proxy profile. While the same TCP stack should negotiate correctly and independently with both clients and servers, it may be desirable to have a TCP stack tuned to servers, such as disabling Nagle's algorithm on the server side but enabling it on the client side. This option only applies to virtual services set for TCP proxy.
- **Auto Gateway**: Return packets are sent to the source MAC address that is associated with the connection instead of returning client data via the default gateway of Avi Vantage. If Vantage has the wrong default gateway, no configured gateway, or multiple gateways, client-initiated return traffic will still flow correctly. The Vantage default gateway will still be used for management and outbound-initiated traffic.
- **Host Name Translation**: If the host header name in a client HTTP request is not the same as this field, or if it is an IP address, Vantage will translate the host header to this name prior to sending the request to a server. If a server issues a redirect with the translated name, or with its own IP address, the redirect's location header will be replaced with the client's original requested host name. Host name translation does not rewrite cookie domains or absolute links that might be embedded within the HTML page. This option is applicable to HTTP virtual services only. This capability may be manually created using HTTP request and response policies.

In this example, the Host Name Translation field is set to "b.com"

	Client		*Avi*		***Server***
==>	*Request sent*	a.com	==>	b.com	*Request received*
<==	*Response received*	a.com/foo	<==	b.com/foo	*Redirect response*

- **SE Group**: The Service Engine group is a container of service engines that includes settings such as high availability properties. Placing a virtual service in a specific Service Engine group may be used to guarantee resource reservation and data plane isolation, such as separating production from test environments. This field may be hidden based on configured Roles or Tenant options. To change an existing virtual service's Service Engine group, first

disable the virtual service, save, switch it to another group, save, then enable and save. Switching a virtual service to another SE group is disruptive to existing client connections.

Virtual Service Scaling

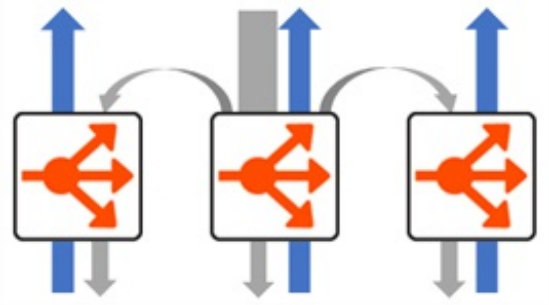
This article covers the following virtual service optimization topics:

- Scaling out a virtual service to an additional Service Engine (SE)
- Scaling in a virtual service back to fewer SEs
- Migrating a virtual service from one SE to another SE

Avi Vantage supports scaling virtual services, which distributes the virtual service workload across multiple SEs to provide increased capacity on demand, thus extending the throughput capacity of the virtual service and increasing the level of high availability.

- Scaling out a virtual service distributes that virtual service to an additional SE if traffic increases. By default, Avi Vantage supports a maximum of four SEs per virtual service.
- Scaling in a virtual service reduces the number of SEs consumed by that virtual service if traffic decreases. A virtual service will always require a minimum of one SE.

Operational Notes



The descriptions above to each of the supported infrastructures (cloud types). This section provides additional information for specific infrastructures.

How Scaling Operates in VMware / OpenStack with Nuage Deployments

For VMware deployments and OpenStack deployments with Nuage, the scaled out traffic behaves as follows:

- The Virtual Service IP is GARPed by the Primary SE. All inbound traffic from clients will arrive on this SE.
- The Primary may handle a percentage of traffic, which is handled normally.
- Excess traffic is forwarded at layer 2 to the MAC address of the additional Service Engines.
- The scaled out traffic to the secondary SEs is processed as normal. The SEs will change the

source IP address of the connection to their own IP address within the server network.

- The servers will respond back to the source IP address of the traffic, which may the primary or one of the secondary SEs.
- The secondary SEs will forward the response traffic directly back to the origin client, bypassing the primary SE.

How Scaling Operates in OpenStack with Neutron Deployments

For OpenStack deployments with native Neutron, server response traffic that is sent to the secondary SEs will be forwarded back through the primary SE before returning to the origin client.

Avi Vantage will issue an Alert if the average CPU utilization of an SE exceeds the designated limit during a five-minute polling period. Alerts for additional thresholds can be configured for a virtual service. The process of scaling in or scaling out must be initiated by an administrator. The CPU Threshold field of the SE Group > High Availability tab defines the minimum and maximum CPU percentages.

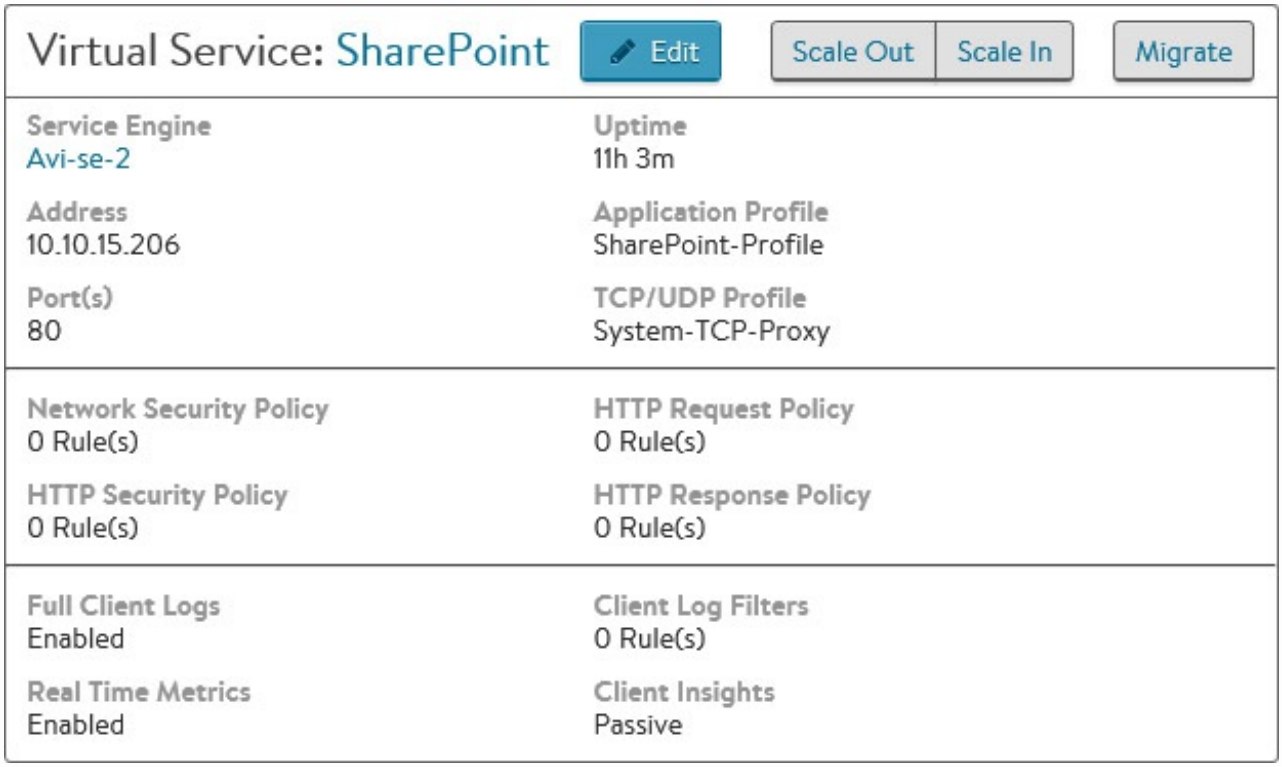
How High Availability Affects Scaling

The High Availability settings within the Service Engine Group for the Virtual Service will determine how and when scaling should be performed.

Scaling Process

The process used to scale out will depend on the level of access, write access or read/no Access, that Avi Vantage has to the hypervisor orchestrator:

- If Avi Vantage is in Write Access mode with write privileges to the virtualization orchestrator, then Avi Vantage will automatically create additional Service Engines when required to share the load. If the Controller runs into an issue when creating a new Service Engine, it will wait a few minutes and then try again on a different host. When AutoScale is enabled, the original Service Engine still owns and ARPs for the Virtual Service IP address to process as much traffic as it can. Some percentage of traffic will be forwarded via layer 2 to the additional Service Engines. When traffic decreases, the Virtual Service automatically scales in back to the original, primary Service Engine.
- If Avi Vantage is in Read Access or No Access mode, an administrator must manually create and configure new Service Engines in the virtualization orchestrator. The Virtual Service can only be scaled out once the Service Engine is both properly configured for the network and connected to the Avi Vantage Controller.



Scaling Out

To manually scale a Virtual Service out when Avi Vantage is operating in Write Access mode:

1. Open the Virtual Service Details page for the Virtual Service that you want to scale.
2. Hover the cursor over the name of the Virtual Service to open the Virtual Service Quick Info popup.
3. Click the Scale Out button to scale the Virtual Service out to an additional Service Engine per click, up to a maximum of four Service Engines.
4. If available, Avi Vantage will attempt to use an existing Service Engine. If none is available or matches reachability criteria, it may create a new SE.
5. In some environments, Avi Vantage may prompt for additional information in order to create a new Service Engine, such as additional IP addresses.

The prompt “Currently scaling out” displays the progress while the operation is taking place.

Scaling out a Virtual Service may take anywhere from a few seconds to a few minutes. The scale out timing depends whether an additional Service Engine exists or if a new one needs to be created, as well as network and disk speeds if creating a new SE.

Scaling In

To manually scale in a Virtual Service in when Avi Vantage is operating in Write Access mode:

1. Open the Virtual Service Details page for the Virtual Service that you want to scale.
2. Hover the cursor over the name of the Virtual Service to open the Virtual Service Quick Info popup.
3. Click the Scale In button to open the Scale In popup window.
4. Select Service Engine to scale in. In other words, which SE should be removed from supporting this Virtual Service.
5. Scale the Virtual Service in by one Service Engine per SE selection, down to a minimum of one Service Engine.

The prompt “Currently scaling in” displays the progress while the operation is taking place.

Migrating

The Migrate option allows graceful migration from one Service Engine to another. During this process, the primary SE will scale out to the new SE and begin sending it new connections. After thirty seconds, the old SE will be deprovisioned from supporting the Virtual Service.

Virtual Service Policies

Policies allow advanced customization of network layer security, HTTP security, HTTP requests, and HTTP responses. A policy may be used to control security, client request attributes, or server

response attributes. Policies are comprised of *matches* and *actions*, similar to an if/then logic. If something is true, then it matches the policy, therefore, perform the following action.

Policies are comprised of one or more rules, which are match/action pairs. A rule may contain many matches, or have many actions. Multiple policies may be configured for a virtual service. Policies may alter the default behavior of the virtual service, or if matching criteria is not met, they may be benign for a particular connection, request, or response.

Policies are not shared; they are defined on a per virtual service basis. While powerful, policies are intended to be simple point and click functionality.

For more advanced capabilities, see DataScripts.

Policies are configured within the policy tab of the virtual service edit screen.

Prioritizing Policies

Policies may be used to recreate similar functionality found elsewhere within Vantage. For instance, a policy may be configured to generate an HTTP redirect from HTTP to HTTPS. This same functionality may be configured within the Secure-HTTP application profile. Because a policy is more specific than a general purpose profile, the policy will take precedence.

If the profile is set to redirect HTTP to HTTPS via port 443, and the policy is set to redirect HTTP to HTTPS on port 8443, the client will be sent to port 8443. (See Execution Priority for more on this topic.)

A virtual service may have multiple policies defined for each of the four types. When multiple policies are defined, they are implemented in the following order of priority.

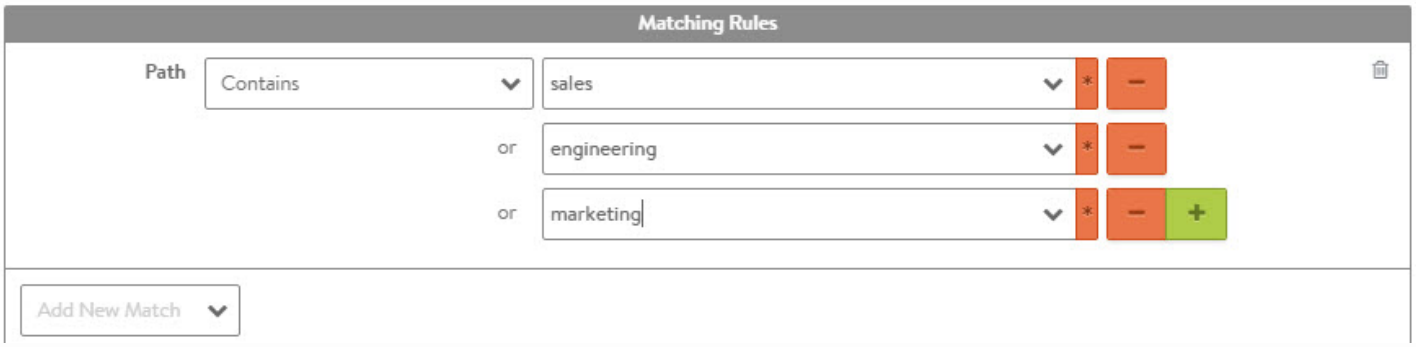
1. Network security policy
2. HTTP security policy
3. HTTP request policy
4. HTTP response policy

For example, a network policy that is set to discard traffic takes precedence over an HTTP request policy to modify a header. Since the connection is discarded, the HTTP request policy will not execute. Each policy type may contain multiple rules, which can be prioritized to process in a specified order. This is done by moving the policies up or down in the ordered list within the Vantage web interface.

Match / Action

All policies are made up of match and action rules, which are similar in concept to “if / then” logic. Administrators set match criteria for all connections, requests, or response to the virtual service.

Vantage then executes the configured actions for all traffic that meets the match criteria.



A single match with multiple entries is treated as “or” operation. For example, if a single match type has the criteria “marketing”, “sales”, and “engineering” set, then the match is true if the path contains “marketing”, or “sales”, or “engineering”.

If a rule has multiple matches configured, then all match types must be true for the action to be executed. In the figure above, both the path and HTTP method matches must be true. Within each of these two match types, only one of the entries to be true for that match type to be true. For HTTP method, a client request must be of type GET or HEAD. Multiple rules may be configured for each policy, and they may be configured to occur in a specified order. If no match is applied, the condition is automatically met and the actions will execute for each connection as appropriate for the policy type.

Matches against HTTP content are case insensitive. This is true for header names and values, cookies, host names, paths, and queries. For HTTP policies, Vantage compares Uniform Resource Identifier (URI) matches against the decoded HTTP URI. Many browsers and web servers encode human-readable format content differently. For example, a browser’s URI encoding might translate the dollar character “$” to “%24”. The Service Engine (SE) translates the “%24” back to “$” before evaluating it against the match criteria.

Create a Policy

The Virtual Service Create/Edit > Rules > Policies tab defines policies that consist of one or more rules that control the flow of requests through the virtual service. To create a policy:

1. **Policy Type:** First select the policy type to add by selecting one of the following categories:

- **HTTP Security:** HTTP security policies perform defined actions such as allow/deny, redirect to HTTPS, or respond with a static page.
- **HTTP Request:** HTTP request policies allow manipulation of HTTP Requests and content switching; they allow customized actions based on client HTTP requests.
- **HTTP Response:** HTTP response policies evaluate responses from the server, and can be used to modify the server's response headers. HTTP response policies are most often used in conjunction with HTTP request policies to provide an Apache Mod_ProxyPass capability for rewriting a website's name between the client and server.
- **Network Security:** Is configured to explicitly allow or block traffic of user-defined types onto the network.

2. **Create Rule:** Create a new rule by clicking the green add rule button and then entering the following information for the new rule:
   - **Enable/Disable:** By default, the new rule will be enabled. The green icon can be clicked to change to gray, which means this rule will be disabled, and will have no effect on traffic while in this state.
   - **Rule Name:** Enter a unique name for the rule in the rule name field, or leave the default system generated name in place.
   - **Logging:** Select the checkbox if you want logging enabled for this rule. When enabled, a log will be generated for any connection or request that matches the rule's match criteria. If a virtual service is already set to log all connections or requests, this logging checkbox will not create a duplicate log. Client logs will be flagged with an entry for the policy type and rule name that matched. When viewing the policy's logs within the logs tab of the virtual service, the logs will be part of the Significant logs option unless the specific connection or request is an error, in which case it may be visible via the default Non-Significant logs filter.
   - **Match:** Add one or more matches using the pull-down menu. The match options will vary depending on the context defined by the policy type you are creating. If a rule is not given a match, then all connections or requests will be considered true or matched.
   - **Action:** Add one or more actions to be taken when the matches are true. The available options will vary depending on the type of rule you are creating.
   - **Save:** Click the save rule button to save the new rule.
3. **Ordering:** Rules are enforced in the order in which they appear in the list. For example, if you add a rule to close a connection based on a client IP address followed by a rule that redirects

an HTTP request from that IP address to a secure HTTP (HTTPS) connection, then Avi Vantage will close the connection without forwarding the request. Alter the order in which rules are applied by clicking the up and down arrow icons until the rules are in the desired order.

Network Security

The following table lists both the available Network Security Match Criteria and the configurable Actions that can occur when a match is made.

Match	**Client IP:** Client IP address or a Group of client addresses. • Use a "-" to specify a range: `10.0.0.0-10.1.255.255` • Use a "/" to specify a netmask: `10.0.0.0/24`
	Service Port: The ports the Virtual Service is listening on.
Actions	**Logging:** Selecting the logging checkbox causes Vantage to log when an action has been invoked.
	Allow / Deny: Explicitly allow or deny any matched traffic. Denied traffic will be issued a reset (RST), unless the system is under a volumetric or denial of service attack, in which case the connection may be silently discarded.
	Rate Limit: Restrict clients from opening greater than the specified number of connections per second in the Maximum Rate. Clients that exceed this number will have their excessive connection attempts silently discarded. If Burst Size is enabled, clients may be able to burst above the Max Rate provided they have not recently been opening connections. This feature may be applied to TCP or UDP. All clients that match the Match criteria will be treated as one bucket. For instance, if no Match is defined, any and all IP addresses will increment the Max Rate counter. Throttling will occur for all new connecting clients. To enable per client throttling, see the Advanced tab for the virtual service. The manual for this page also contains a more robust description of connecting throttling.

HTTP Security

The following table lists both the available HTTP Security Match Criteria and the configurable Actions that can occur when a match is made.

Match	**Client IP:** Client IP address or a Group of client addresses. • Use a "-" to specify a range: `10.0.0.0-10.1.255.255` • Use a "/" to specify a netmask: <code<10.0.0.0 24 </code

	Service Port: The ports the Virtual Service is listening on.
	Protocol Type: HTTP or HTTPS. Example: ***https://***www.avinetworks.com/marketing/index.html?a=1&b=2
	HTTP Method: The method used by the client request. The match is true if any one of the methods that an administrator specifies is true.
	HTTP Version: True if the client version is .9, 1.0, or 1.1
	Path: The path or a group of paths. Paths do not need to begin with a forward slash (/). For comparison purposes, Vantage automatically omits any initial slash specified in the match field. Example: https://www.avinetworks.com/**marketing/index.html**?a=1&b=2
	Query: A query or a group of queries. Do not add the leading '?' or '&' characters to a match. Example: *https://www.avinetworks.com/marketing/index.html?**a=1&b=2***
	Headers: True if a header exists, or if it exists and contains a specified value
	Cookie: True if a cookie exists, or if it exists and contains a specified value
	Host Header: The request's host header. Example: *https://**www.avinetworks.com**/marketing/index.html?a=1&b=2*
	Location Header: The Location Header may not exist for every website.
	HTTP Status: The status of the response, such as 200 (success), 404 (file not found), or similar. The statuses can be separated by commas, or be a range. For example: 301, 302, 307, 308, 300-599
	Response Header: Match based on a specific header sent by the server.
Actions	**Logging:** Selecting the logging checkbox causes Vantage to log when an action has been invoked.
	Action Allow: Allows matched requests to continue on to further policies or to the destination pool servers.
	Action Close Conn: Matched requests will cause Vantage to close the TCP connection that received the request via a FIN. Many browsers open multiple connections, which are not closed unless requests sent over those connections also trigger a close connection action.
	Redirect To HTTPS: Respond to the request with a temporary redirect to the desired port for SSL.

Action Send Response: Vantage may serve an HTTP response using HTTP status code 200 (success), 403 (unauthorized), or 404 (file not found). A default page is rendered by the browser for each of these status codes, or you may upload a custom .html file. This file may have links to images or other files, but only the initial html file will be stored and served via the Send Response.

Policy Tokens

In more complex scenarios, an administrator may wish to capture data from one location and apply it to another location. Vantage supports the use of variables and tokens, which can be used for this purpose.

Variables may be used to insert dynamic data into the Modify Header Actions of HTTP Request and HTTP Response policies. Two variables are supported, $client_ip and $vs_port. For example, a new header may be added to an HTTP request called origin_ip, with a value set to $client_ip, which will insert the client's source address as the value of the header.

Tokens may be used to find and reorder specific parts of the HTTP hostname or path. For example, it is possible to rewrite support.avinetworks.com to www.avinetworks.com/support. Tokens can be used for HTTP host and HTTP path. The tokens are derived from the original URL. Token delimiter in host header is "." and in the URL path is "/".

Request URL:	www	avi	com	support	docs	index.htm
Token:	host[0]	host[1]	host[2]	path[0]	path[1]	path[2]

In the example above, the client request is broken down into HTTP host and HTTP path. Each section of the host and path are further broken down according to the "." and "/" delimiters for host and path. A host or path token may be used in an Action to rewrite a header, a host, or a path. In the example below, the redirect will send requests to documentation.avinetworks.com.

In addition to using the host[1], host[2], host[3] convention, a colon may be used to denote the system should continue till the end of the host or path. For instance, host[1:] implies to use avinetworks, followed by any further host fields. The result would be avinetworks.com. This is

especially useful in a path, which may contain many levels. Tokens may also specify a range, such as path[2:5]. Host and path tokens may also be abbreviated as “h” and “p”, such as h[1:] and p[2][2].

In the rewrite URL, redirect, and rewrite location header actions, the host component of the URL can be specified in terms of tokens - the tokens can be constants strings, tokens from existing host and path component of the URL.

e l p a x E

Request URL:	paris	france	avi	com	region	index.htm	
Token:	host[0]	host[1]	host[2]	host[3]	path[0]	path[1]	
New Host:	path[0].host[2:]						
New Path:	/host[1]/host[0]/path[1]						
New URL:	region.avi.com/france/paris/index.htm						

e l p a x E

Request URL:	www1	avi	com	sales	foo	index.htm	auth=true
Token:	host[0]	host[1]	host[2]	path[0]	path[1]	path[2]	(query)
New Host:	www.host[1:]						
New Path:	/host[0]/path[0:]						
Query:	Keep Query enabled						
New URL:	www.avi.com/www1/sales/foo/index.htm?auth=true						

- If the host header contains an IPv4 address and not a FQDN, and the rewrite URL or redirect action refers to a host token (e.g. host[0], host[1,2], etc.) the rule action is skipped and the next rule is evaluated.
- If the host header or path contains less tokens than that referenced in the action, then the rule action is skipped. For example, if the host name in host header has only 3 tokens (host name www.avinetworks.com - token host[0] = www, host[1][1] = avinetworks, host[2][2] = com). If the action refers to host[4] the rule action is skipped.
- If the location header in the HTTP response contains an IPv4 address and the response

policy action is rewrite location header which refers to host tokens, the rule action is skipped.

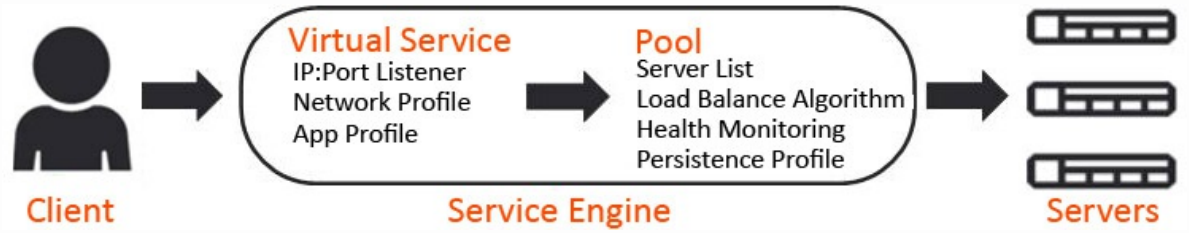
- Rule engine does not recognize octal or hexadecimal IPv4 address in the host address. That is, the rule action will not be skipped if the host header has octal/hexadecimal IPv4 address and the action references a host token such as host[1][1], etc.
- If an HTTP Request arrives with multiple host headers, the first host header will be used.
- Per RFC, HTTP 1.1 Requests must have a non-empty host header, else a 400 'Bad Request' HTTP Response will be returned by Vantage.
- The HTTP processing is performed against decoded URIs.

Pools

This chapter contains the following sections:

- Pool Page
- Pool Analytics Page
- Pool Logs Page
- Pool Health Page
- Pool Servers Page
- Pool Events Page
- Pool Alerts Page

Pools maintain the list of servers assigned to them and perform health monitoring, load balancing, persistence, and functions that involve Vantage-to-server interaction. A typical virtual service will point to one pool; however, more advanced configurations may have a virtual service content switching across multiple pools via HTTP request policies or DataScripts. A pool may only be used or referenced by only one virtual service at a time.



Creating a virtual service using the basic method automatically creates a new pool for that virtual service, using the name of the virtual service with a *-pool* appended. When creating a virtual service via the advanced mode, an existing, unused pool may be specified, or a new pool may be created.

Pools Page

Select *Applications* > *Pools* to open the pools page. This page displays a high level overview of configured pools. This page includes the following functions:

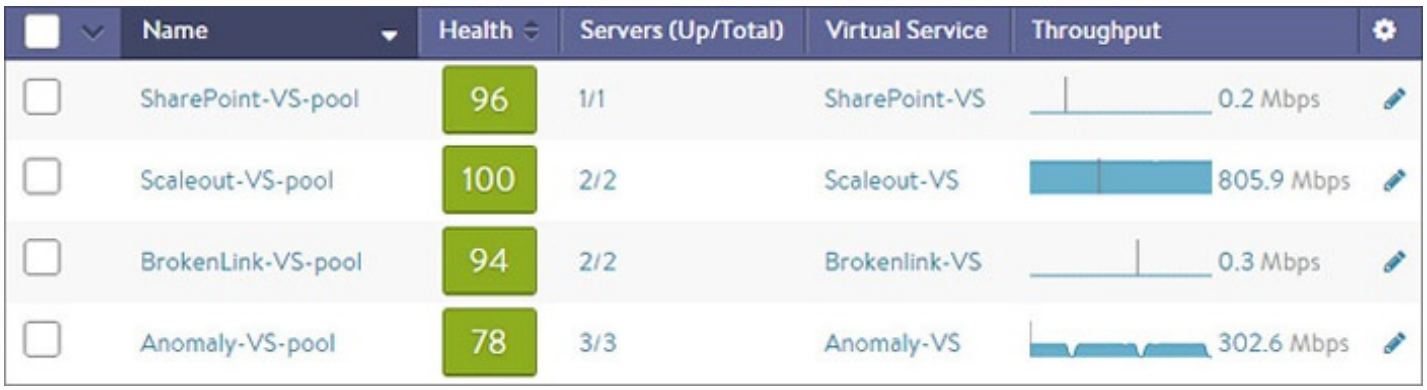
Search: Filter the list of pools by entering full or partial name of a pool.

Create: Opens the create pool popup.

Edit: Opens the edit pool popup.

Delete: Select one or more pools in the table and click the delete button at the top left of the table to delete the pools. Only unused pools (with a gray health score) may be deleted. Pools that are in use (e.g. pools referenced by a virtual service) must first be disassociated from the virtual service by deleting or editing the VS.

The table on this page displays the following information for each pool. The columns shown may be modified via the sprocket icon in the top right of the table:



- **Name**: Lists the name of each pool. Clicking the name opens the Analytics tab of the Pool Details page.
- **Health**: Provides both a number from 1-100 and a color-coded status to provide quick information about the health of each pool. This will be gray if the pool is unused, such as not associated with a virtual service or associated with a VS that can not or has not been placed on a Service Engine.
  - Hovering the cursor over the health score opens the pool's Health Score popup.
  - Clicking the View Insights link at the bottom of the pool's Health Score popup opens the health Insights tab of the Pool Detail page.
  - Clicking elsewhere within the pool's Health Score popup opens the Analytics tab of the Pool Details page.
- **Servers**: Displays the number of servers in the pool that are up out of the total number of servers assigned to the pool. For example, *2/3* indicates that two of the three servers in the

pool are successfully passing health checks and are considered up.

- **Virtual Service**: The VS the pool is assigned to. Clicking a name in this column opens the VS Analytics tab of the Virtual Service Details page. If no virtual service is listed, this pool is considered unused.
- **Throughput**: Thumbnail chart of the throughput in Mbps for each pool for the time frame selected.
  - Hovering the cursor over this graph shows the throughput at the selected time.
  - Clicking a graph opens the Analytics tab of the pool's Details page.

Pool Details Page

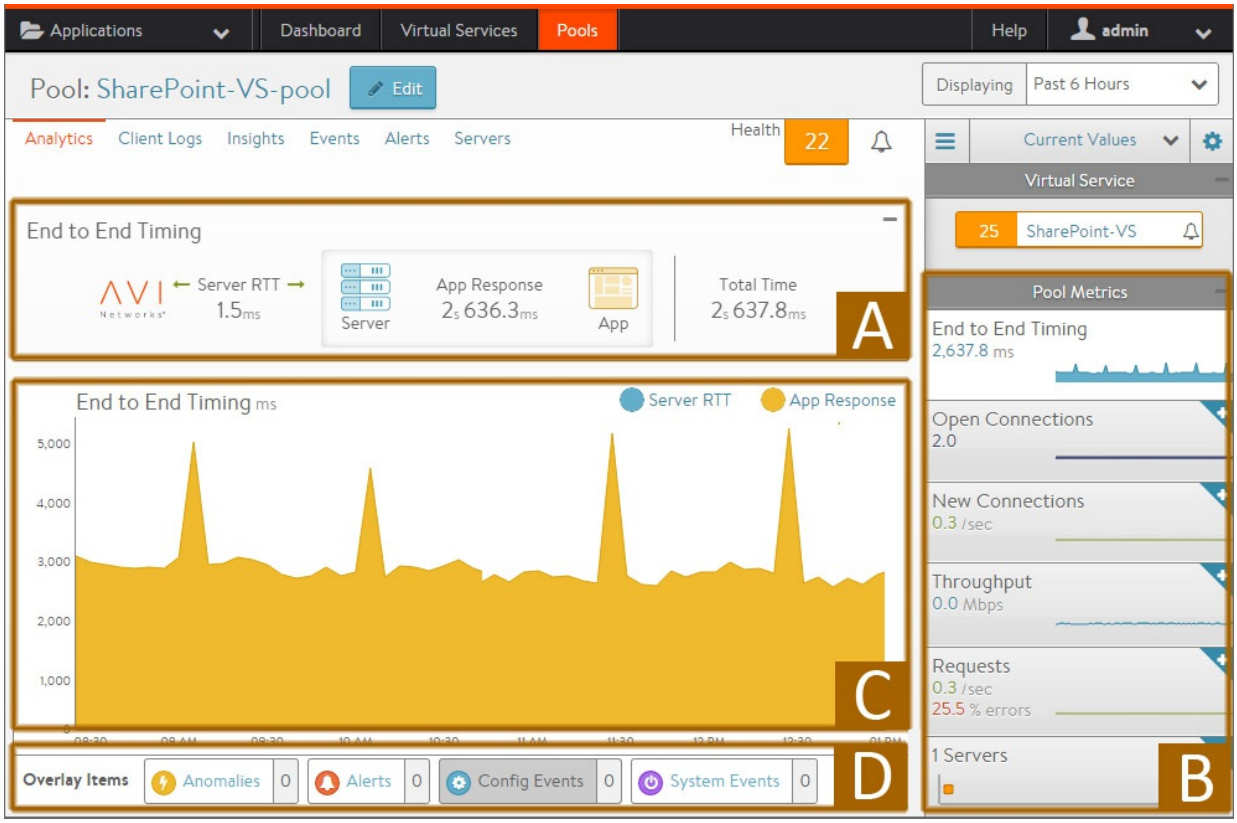
Clicking into a pool brings up the Details pages, which provide deeper views into the current pool.

This page contains the following sub-pages:

- Analytics
- Logs
- Health
- Servers
- Events
- Alerts

Pool Analytics Page

The pool's Analytics tab presents information about various pool performance metrics. Data shown is filtered by the time period selected.



See the following for detailed information about this tab:

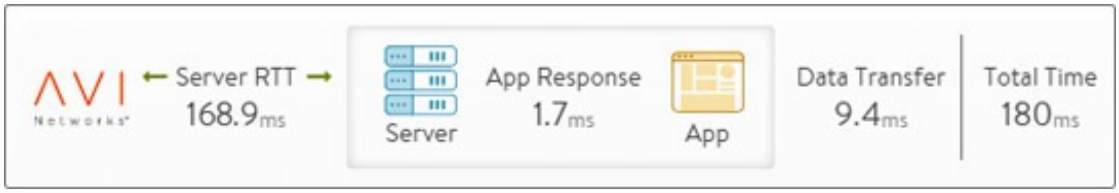
- (A) End-to-End Timing
- (B) Metric Tiles
- (C) Chart Pane
- (D) Overlays Pane
  - Anomalies
  - Alerts
  - Config Events
  - System Events

Pool End-to-End Timing

The End to End Timing pane at the top of the Analytics tab of the Pool Details Page provides a high-level overview of the quality of the end-user experience and where any slowdowns may be occurring. The chart breaks down the time required to complete a single transaction, such an HTTP request.

It may be helpful to compare the end-to-end time against other metrics, such as throughput, to see how increases in traffic impact the ability of the application to respond. For instance, if new

connections double but the end-to-end time quadruples, you may need to consider adding additional servers.



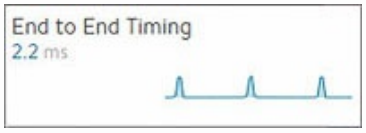
From left to right, this pane displays the following timing information:

- **Server RTT:** This is Service Engine to server round trip latency. An abnormally high Server RTT may indicate either that the network is saturated or more likely that a server's TCP stack is overwhelmed and cannot quickly establish new connections.
- **App Response:** The time the servers take to respond. This includes the time the server took to generate content, potentially fetch back-end database queries or remote calls to other applications, and begin transferring the response back to Vantage. This time is calculated by subtracting the Server RTT from the time of the first byte of a response from the server. If the application consists of multiple tiers (such as web, applications, and database), then the App Response represents the combined time before the server in the pool began responding. This metric is only available for a layer 7 virtual service.
- **Data Transfer:** Data Transfer represents the average time required for the server to transmit the requested file. This is calculated by measuring from the time the Service Engine received the first byte of the server response until the client has received the last byte, which is measured as the when the last byte was sent from the Service Engine plus one half of a client round trip time. This number may vary greatly depending on the size of objects requested and the latency of the server network. The larger the file, the more TCP round trip times are required due to ACKs, which are directly impacted by the Client RTT and Server RTT. This metric is only used for a Layer 7 virtual service.
- **Total Time:** Total time from when a client sent a request until they receive the response. This is the most important end-to-end timing number to watch, because it is the sum of the other four metrics. As long as it is consistently low, the application is probably successfully serving traffic.

Pool Metrics

The sidebar metrics tiles contain the following metrics for the pool. Clicking any metric tile will change the main chart pane to show the chosen metric.

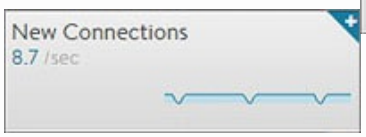
- **End to End Timing:** Shows the total time from the pool's End to End Timing graph. To see the complete end-to-end timing, including



the client latency, see the Analytics tab of the Virtual Service Details page, which includes the client to Service Engine metric.

- **Open Connections:** The number of open (existing) connections during the selected time period.

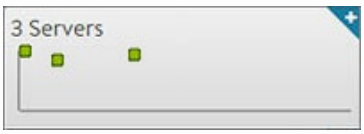
- **New Connections:** The number of client connections that were completed or closed over the selected time period. See this article for an explanation of new versus closed connections per second.



- **Throughput:** Total bandwidth passing between the virtual service and the servers assigned to the pool. This throughput number may be different than the virtual service throughput, which measures throughput between the client and the virtual service. Many features may affect these numbers between the client and server side of Vantage, such as caching, compression, SSL, and TCP multiplexing. Hovering your mouse cursor over this graph displays the throughput in Mbps for the selected time period.

- **Requests:** The number of HTTP requests sent to the servers assigned to the pool. This metric also shows errors sent to servers or returned by servers. Any client requests that received an error generated by Vantage as a response (such as a 500 when no servers are available) are not be forwarded to the pool and will not be tracked in this view.

- **Servers:** Displays the number of servers in the pool and their health. The X-axis represents the number of HTTP requests or connections to the server, while the Y-axis represents the health score of the server. The chart allows you to view servers in relation to their peers within the pool, thus helping to spot outliers. Within the chart pane, click and drag the mouse over server dots to select and display a table of the highlighted servers below the Chart pane. The table provides more details about these servers, such as hostname, IP address, health, new connections or requests, health score, and the server's static load balanced ratio. Clicking on the name of a server will jump to the pool's Server Insight page, which shows additional health and resource status.



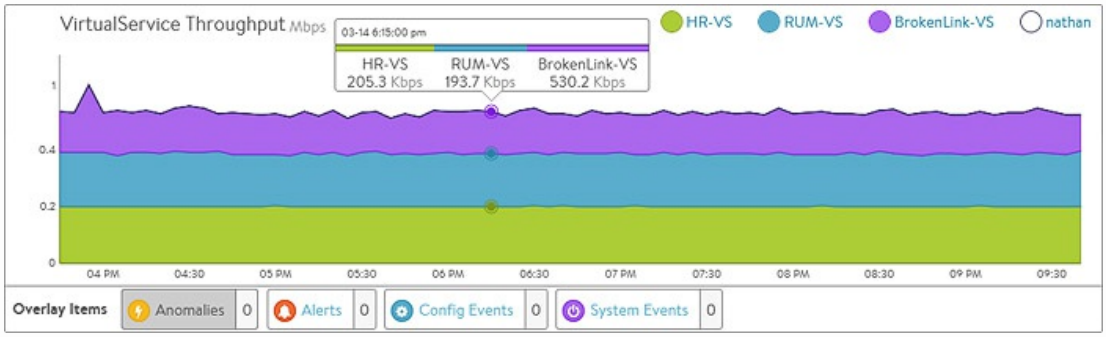
Pool Chart Pane

The main chart pane in the middle of the Analytics tab displays a detailed historical chart of the selected metric tile for the current pool.

- Hovering the mouse over any point in the chart will display the results for that selected time in a popup window.
- Clicking within the chart will freeze the popup at that point in time. This may be useful when

the chart is scrolling as the display updates over time.

- Clicking again will unfreeze the highlighted point in time.



Many charts contain radio buttons in the top right that allow customization of data that should be included or excluded from the chart. For example, if the End to End Timing chart is heavily skewed by one very large metric, then deselecting that metric by clearing the appropriate radio button will re-factor the chart based on the remaining metrics shown. This may change the value of the vertical Y-axis.

Some charts also contain overlay items, which will appear as color-coded icons along the bottom of the chart.

Pool Overlays Pane

The overlays pane is used to highlight important events within the timeline of the chart pane. This feature helps correlate anomalies, alerts, configuration changes, or system events with changes in traffic patterns.

Within the overlays pane:

- Each overlay type displays the number of entries for the selected time period.
- Clicking an overlay button toggles that overlay's icons in the chart pane. The button lists the number of instances (if any) of that event type within the selected time period.
- Selecting an overlay button displays the icon for the selected event type along the bottom of the chart pane. Multiple overlay icon types may overlap. Clicking the overlay type's icon in the chart pane will bring up additional data below the overlay Items bar. The following overlay types are available:
  - **Anomalies**: Display anomalous traffic events, such as a spike in server response time, along with corresponding metrics collected during that time period.
  - **Alerts**: Display alerts, which are filtered system-level events that have been deemed important enough to notify an administrator.
  - **Config Events**: Display configuration events, which track configuration changes made to

Vantage by either an administrator or an automated process.

- **System Events**: Display system events, which are raw data points or metrics of interest. System Events can be noisy, and are best used as the basis of alerts which filter and classify raw events by severity.

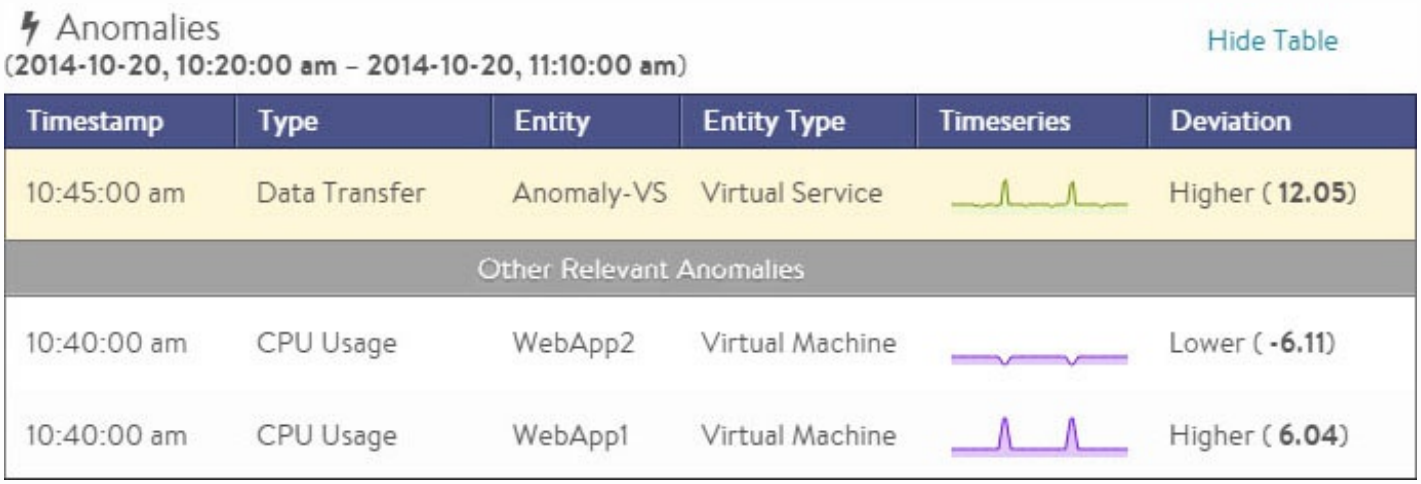
Pool Anomalies Overlay

The anomalies overlay displays periods during which traffic behavior was considered abnormal based on recent historical moving averages. Changing the time interval will provide greater granularity and potentially show more anomalies.

Clicking the Anomalies Overlay button displays yellow anomaly icons in the chart pane. Selecting one of these icons within the chart pane brings up additional information in a table at the bottom of the page. During times of anomalous traffic, Vantage records any metrics that have deviated from the norm, which may provide hints as to the root cause of the anomaly.

An anomaly is defined as a metric that has a deviation of 4 sigma or greater across the moving average of the chart.

Anomalies are not recorded or displayed while viewing with the Real Time display period.



- **Timestamp:** Date and time when the anomaly was detected. This may either span the full duration of the anomaly, or merely be near the same time window.
- **Type:** The specific metric deviating from the norm during the anomaly period. To be included, the metric deviation must be greater than 4 sigma. Numerous types of metrics, such as CPU utilization, bandwidth, or disk I/O may trigger anomalous events.
- **Entity:** Name of the specific object that is reporting this metric.

- **Entity Type:** Type of entity that caused the anomaly. This may be one of the following:
  - Virtual Machine (server); these metrics require Vantage to be configured for either read or write access to the virtualization orchestrator such as vCenter or OpenStack. In the example shown above, CPU utilization of the two servers was learned by querying vCenter.
  - Virtual service
  - Service Engine
- **Time Series:** Thumbnail historical graph for the selected metric, including the most current value for the metric which will be data on the far right. Moving the mouse over the chart pane will show the value of the metric for the selected time. Use this to compare the normal, current, and anomaly time periods.
- **Deviation**: Change or deviation from the moving average, either higher or lower. The time window for the moving average depends on the time series selected for the Analytics tab.

Pool Alerts Overlay

The alerts overlay displays the results of any events that meet the filtering criteria defined in the alerts tab. Alerts notify administrators about important information or changes to a site that may require immediate attention.

Alerts may be transitory, meaning that they may expire after a defined period of time. For instance, Vantage may generate an alert if a server is down and then allow that alert to expire after a specified time period once the server comes back online. The original event remains available for later troubleshooting purposes.

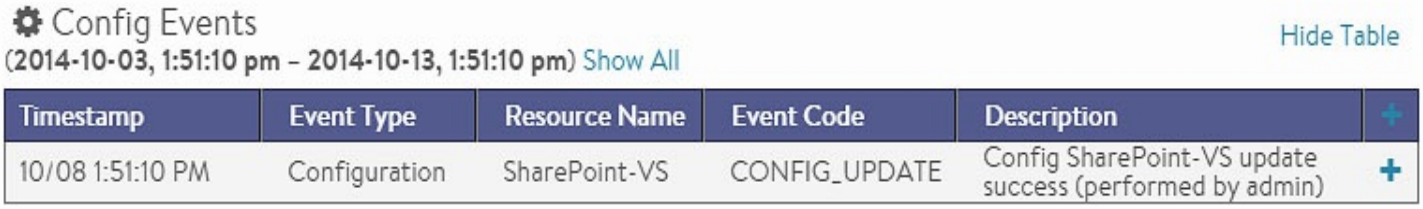
Clicking the alerts icon in the overlay items bar displays any red alerts icons in the chart pane. Selecting one of these chart alerts will bring up additional information below the overlay Items bar, which will show the following information:

- **Timestamp**: Date and time when the alert occurred.
- **Resource Name**: Name of the object that is reporting the alert.

- **Level**: Severity of the alert. You can use the priority level to determine whether additional notifications should occur, such as sending an email to administrators or sending a log to Syslog servers. The level may be one of the following:
  - **High**: Red
  - **Medium**: Yellow
  - **Low**: Blue
- **Summary**: Brief description of the event.
- **Actions**:
  - **Dismiss** the alert with the red**X** to remove it from both the list shown and the alert icon the chart pane. Dismissing an alert here is the same as dismissing it via the bell icon at the top of the screen next to the health score, or dismissing it via the alerts tab.
  - **Edit** the alert filter to make Vantage more or less sensitive to generating new Alerts.
- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for an Alert opens and closes a sub-table showing more detail about the alert. This will typically show the original events that triggered the alert.

Pool Config Events Overlay

The config events overlay displays configuration events, such as changing the Vantage configuration by adding, deleting, or modifying a pool, virtual service, or Service Engine, or an object related to the object being inspected. If traffic dropped off at precisely 10:00am, and at that time an administrator made a change to the virtual services security settings, there's a good chance the cause of the change in traffic was due to the (mis)configuration.



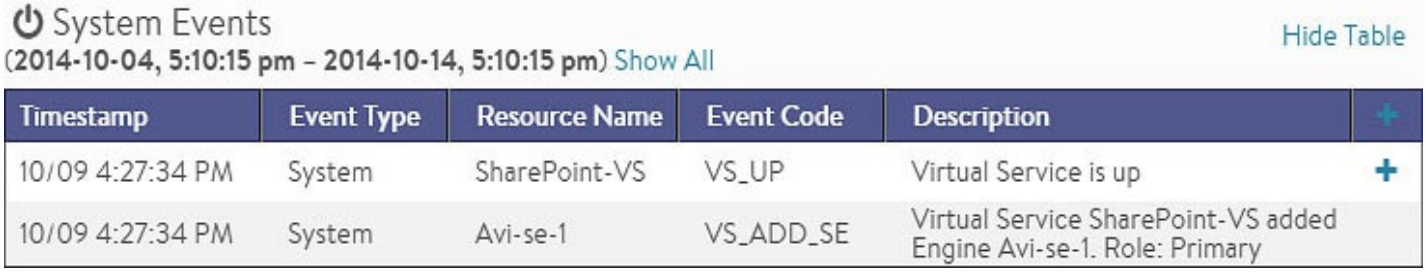
Clicking the Config Events icon in the Overlay Items bar displays any blue config event icons in the chart pane. Selecting one of these chart alerts will bring up additional information below the Overlay Items bar, which will show the following information:

- **Timestamp**: Date and time when the configuration change occurred.
- **Event Type**: This event type will always be scoped to configuration event types.
- **Resource Name**: Name of the object that has been modified.
- **Event Code**: There are three event codes:

    - CONFIG_CREATE
    - CONFIG_UPDATE
    - CONFIG_DELETE
- **Description**: Brief description of the event.
- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for a configuration event either expands or contracts a sub-table showing more detail about the event. When expanded, this shows a difference comparison of the previous configuration versus the new configuration, as follows:
    - Additions to the configuration, such as adding a health monitor, will be highlighted in green in the new configuration.
    - Removing a setting will be highlighted in red in the previous configuration.
    - Changing an existing setting will be highlighted in yellow in both the previous and new configurations.

Pool System Events Overlay

This overlay displays system events relevant to the current object, such as a server changing status from up to down or the health score of a virtual service changing from 50 to 100.



Clicking the system events icon in the overlay items bar displays any purple system event icons in the Chart Pane. Select a system event icon in the chart pane to bring up more information below the overlay items bar.

- **Timestamp**: Date and time when the system even occurred.
- **Event Type**: This will always be system.
- **Resource Name**: Name of the object that triggered the event.
- **Event Code**: High-level definition of the event, such as VS_Health_Change or VS_Up.
- **Description**: Brief description of the system event.
- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for a system event expands or contracts that system event to show more information.

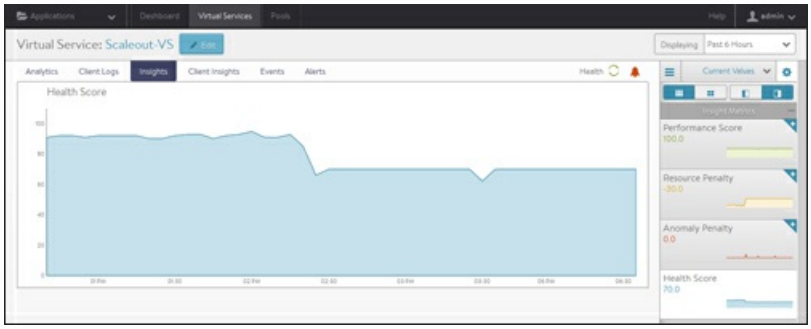
Pool Logs Page

Client logs viewed from within a pool are identical to the logs shown within a virtual service, except they are filtered to only show log data specific to the pool. For instance, information such as End to End Timing is only shown from the Service Engine to the servers, rather than from the clients to the servers. Viewing logs within a pool may be useful when a virtual service is performing content switching across multiple pools. It is still possible within the virtual service logs page to add a filter for a specific pool, which would then provide complete End to End Timing for connections or requests sent to the specified pool.

For the complete descriptions of logs, see theVS logs page help.

Pool Health Page

The health tab presents a detailed breakdown of health score information for the pool.



The health score of a pool is comprised from the following scores:

- **Performance**: Performance score (1-100) for the selected item. A score of 100 is ideal, meaning clients are not receiving errors and connections or requests are quickly returned.
- **Resource Penalty**: Any penalty assessed because of resource availability issues is assigned a score, which is then subtracted from the performance score. A penalty score of 0 is ideal, meaning there are no obvious resource constraints on Vantage or servers.
- **Anomaly Penalty**: Any penalty assessed because of anomalous events is assigned a score, which is then subtracted from the performance score. An ideal score is 0, which means Avi Vantage has not seen recent anomalous traffic patterns that may imply future risk to the site.
- **Health Score**: The final health score for the selected item equals the performance score minus the Resource and anomaly penalty scores

The sidebar tiles show the scores of each of the three subcomponents of the health score, plus the total score. To determine why a pool may have a low health score, select one of the first three tiles that is showing a sub-par score.

This will bring up additional sub-metrics which feed into the top level metric / tile selected. Hover the mouse over a time period in the main chart to see the description of the score degradation. Some tiles may have additional information shown in the main chart section that requires scrolling

down to view.

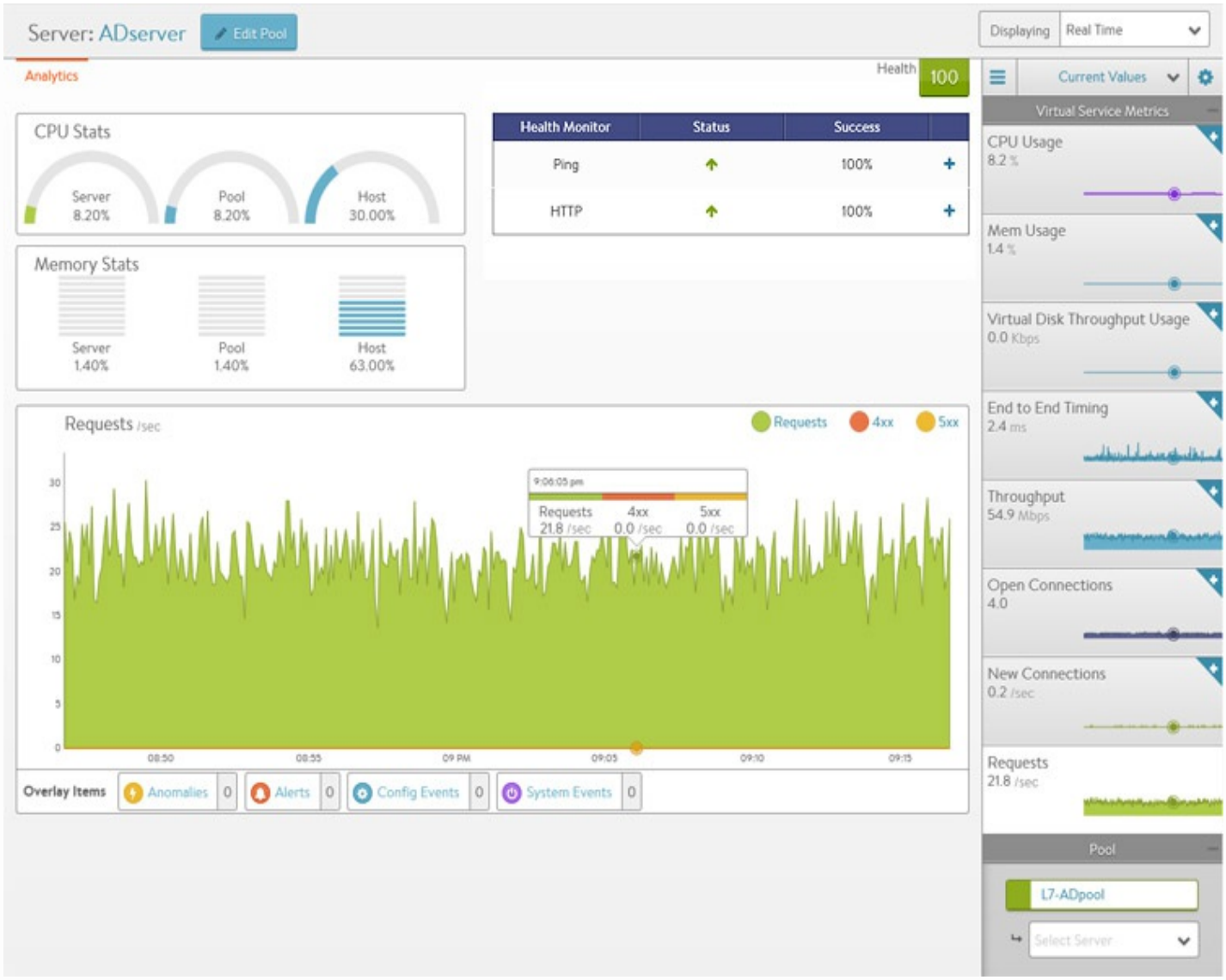
Pool Servers Page

Information for each server within a pool is available within the Server Details Page. This page allows views into correlation between server resources, application traffic, and response times.

Server Page

The Server Page may be accessed by clicking on the server's name from either the Pool > Servers page or the Pool > Analytics Servers tile. When viewing the Server Details page, the server shown is within the context of the pool it was selected within. Rephrased, if the server (IP:Port) is a member of two or more pools, the stats and health monitors shown are only for the server within the context of the viewed pool.

Not all metrics within the Server Page are available in all environments. For instance, servers that are not virtualized or hooked into a hypervisor are not able to have their physical resources displayed.



The statistics can be changed or skewed by switching between Average Values, Peak Values, and Current Values. To see the highest CPU usage over the past day, change the time to 24 hour and the Value to Peak. This will show the highest stats recorded during the past day.

- **CPU Stats:** The CPU Stats box shows the CPU usage for this server, the average during this time period across all servers in the pool, and the hypervisor host.
- **Memory Stats:** The memory Stats box shows the Memory usage for this server, the average during this time period across all servers in the pool, and the hypervisor host.
- **Health Monitor**: This table shows the name of any health monitors configured for the pool. The Status column shows the most current up or down health of the server. The Success column shows the percentage of health monitors that passed or failed during the display time frame. Clicking the plus will expand the table to show more info for a down server. See Why a Server Can Be Marked Down.
- **Main Panel**: The large panel shows the highlighted metric, similar to the Virtual Service

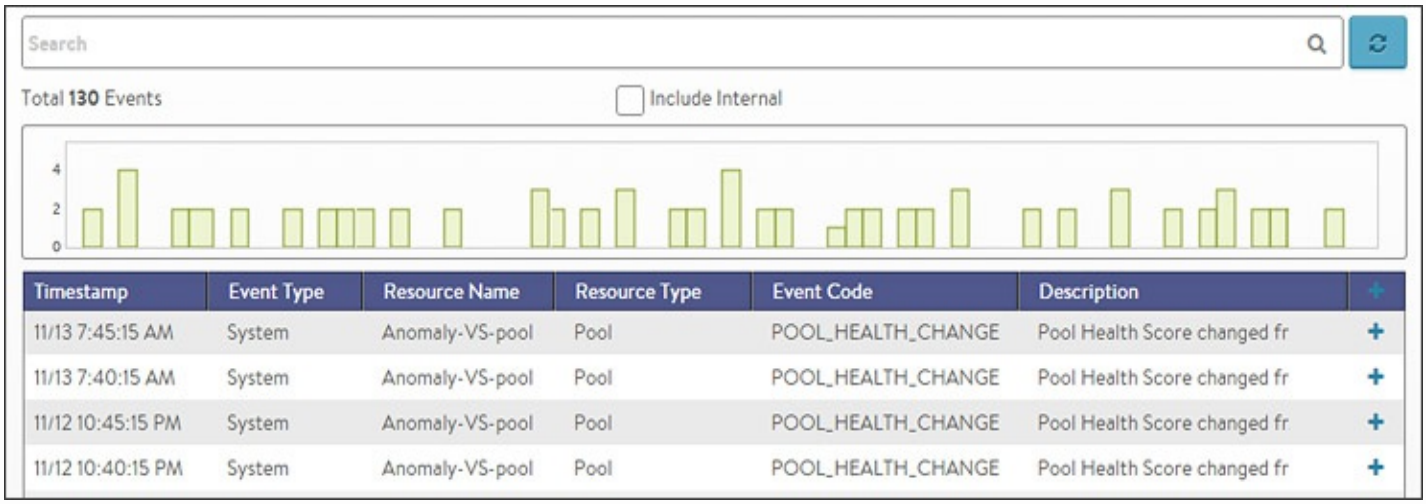
Details and Pool Details pages. Overlay Items shows anomalies, alerts, configuration events, and system events that are related to this server within the pool.

- **Pool Tile Bar**: The pool in the top right bar shows the health of the pool. This can also be used to jump back up to the Pool Page. Under the pool name is a pull-down menu that allows quick access to jumping to the other servers within the pool.

- **Metrics Tile Bar**: The metrics options will vary depending on the hypervisor Vantage is plugged into. For non-virtualized servers, the metrics are limited to non-resource metrics, such as end-to-end timing, throughput, open connections, new connections, and requests. Other metrics that may be shown include CPU, memory, and virtual disk throughput.

Pool Events Page

The events tab presents system-generated events over the time period selected for the pool. System events are applicable to the context in which you are viewing them. For example, when viewing events for a pool, only events that are relevant to that pool are displayed.



The top of this tab displays the following items:

- **Search**: The search field allows you to filter the events using whole words contained within the individual events.
- **Refresh**: Clicking refresh updates the events displayed for the currently-selected time.
- **Number**: The total number of events being displayed. The date/time range of those events appear beneath the search field on the left.
- **Clear Selected**: If filters have been added to the Search field, clicking the Clear Selected (X) icon on the right side of the search bar will remove those filters. Each active search filter will also contain an X that you can click to remove the specific filter.
- **Histogram**: The histogram shows the number of events over the period of time selected. The X-axis is time, while the Y-axis is the number of events during that bar's period of time.
    - Hovering the cursor over a histogram bar displays the number of entries represented by

that bar, or period of time.
    - Click and drag inside the histogram to refine the date/time period which further filters the events shown. When drilling in on the time in the histogram, a zoom to selected link appears above the histogram. This expands the drilled in time to expand to the width of the histogram, and also changes the displaying pull-down menu to custom. To return to the previously selected time period, use the displaying pull-down menu.

The table at the bottom of the events tab displays the events that matched the current time window and any potential filters. The following information appears for each event:

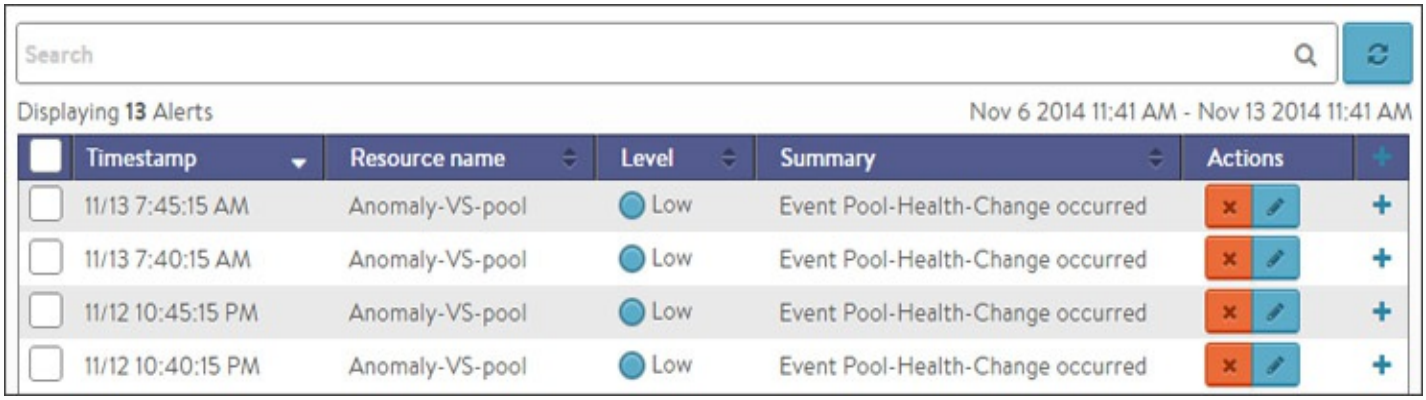
- **Timestamp**: Date and time the event occurred. Highlighting a section of the histogram allows further filtering of events within a smaller time window.
- **Event Type**: This may be one of the following:
    - **System**: System events are generated by Vantage to indicate a potential issue or create an informational record, such as VS_Down,
    - **Configuration**: Configuration events track changes to the Vantage configuration. These changes may be made by an administrator (through the CLI, API, or GUI), or by automated policies.
- **Resource Name**: Name of the object related to the event, such as the pool, virtual service, Service Engine, or Controller.
- **Event Code**: A short event definition, such as Config_Action or Server_Down.
- **Description**: A complete event definition. For configuration events, the description will also show the username that made the change.
- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for an event log either expands or contracts that event log. Clicking the + and – icons in the table header expands and collapses all entries in this tab.

For configuration events, expanding the event displays a difference comparison between the previous and new configurations.

- New fields will appear highlighted in green in the new configuration
- Removed fields will appear highlighted in red.
- Changed fields will show highlighted in yellow

Pool Alerts Page

The alerts tab displays user-specified events for the selected time period. You can configure alert actions and proactive notifications via Syslog or email in the Notifications tab of the Administration page. Alerts act as filters that provide notification for prioritized events or combinations of events through various mechanisms. Avi Vantage includes a number of default alerts based on events deemed to be universally important.



The top of this tab shows the following items:

- **Search**: The search field allows you to filter the alerts using whole words contained within the individual alerts.
- **Refresh**: Clicking refresh updates the alerts displayed for the currently-selected time.
- **Number**: The total number of alerts being displayed. The date/time range of those alerts appear beneath the search field on the left.
- **Dismiss**: Select one or more alerts from the table below then click dismiss to remove the alert from the list.

Alerts are transitory, which means they will eventually and automatically expire. Their intent is to notify an administrator of an issue, rather than being the definitive record for issues. Alerts are based on events, and the parent event will still be in the Events record.

The table at the bottom of the Alerts tab displays the following alert details:

- **Timestamp**: Date and time when the alert was triggered. Changing the time interval using the display pull-down menu may potentially show more alerts.
- **Resource Name**: Name of the object that is the subject of the alert, such as a Server or virtual service.
- **Level**: Severity level of the alert, which can be high, medium, or low. Specific notifications can be set up for the different levels of alerts via the Administration page's Alerts Overlay.
- **Summary**: Summarized description of the alert.
- **Action**: Click the appropriate button to act on the alert:
  - **Dismiss**: Clicking the red X dismisses the alert and removes it from the list of displayed alerts.
  - **Edit**: Clicking the blue pencil icon opens theEdit Alert Config popup for the alert configuration that triggered this alert. This can include a verbose and customized description of the alert or allow an administrator to alter settings such as the severity of the alert.
- **Expand/Contract**: Clicking the plus (+) or minus sign (-) for an event log either expands or contracts that event log to display more information. Clicking the + and – icon in the table

header expands and collapses all entries in this tab

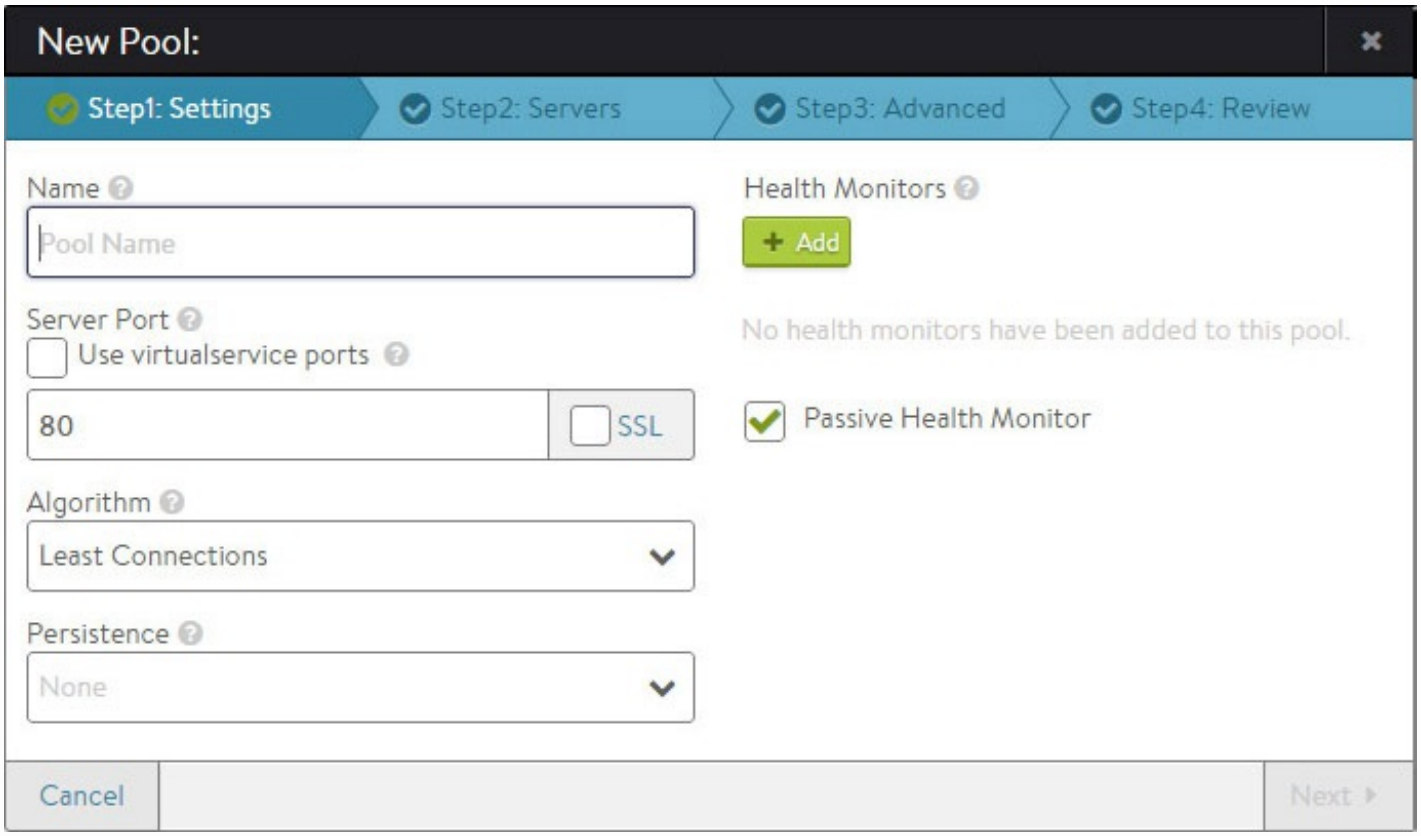
Create Pool

The Create Pool popup and the Edit Pool popup share the same interface that consists of the following tabs:

1. Settings
2. Servers
3. Advanced
4. Review

Create Pool: 1 Settings

The Create/Edit Pool > Settings tab contains the basic settings for the pool. The exact options shown may vary depending on the types of clouds configured in Vantage. For instance, servers in VMware may show an option to “Select Servers by Network” or Cisco ACI integration may show lists of “End Point Groups”.



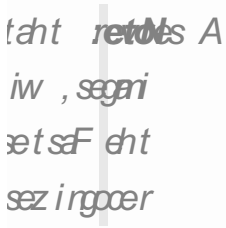
To add or edit Pool settings:

- **Name:** Provide a unique name for the pool.

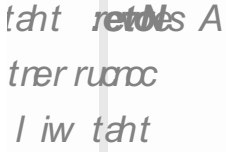
- **Default Server Port:** Select one of the following:
    - **Default Server Port:** New connections to servers will use this destination service port. The default port is 80, unless it is either inherited from the virtual service (if the pool was created during the same workflow), or the port was manually assigned. The default server port setting may be changed on a per-server basis by editing the Service Port field for individual servers in the Servers tab.
    - **SSL:** Enables SSL encryption between the Vantage Service Engine and the back-end servers. This is independent from the SSL option in the virtual service, which enables SSL encryption from the client to the Vantage Service Engine.
        - **SSL Profile:** Determines which SSL versions and ciphers Vantage will support when negotiating SSL with the server.
        - **Server SSL Certificate Validation PKI Profile:** This option validates the certificate presented by the server. When not enabled, the Service Engine automatically accepts the certificate presented by the server when sending health checks. See the PKI Profile section for additional help on certificate validation.
        - **Service Engine Client Certificate:** When establishing an SSL connection with a server, either for normal client-to-server communications or when executing a health monitor, the Service Engine will use this certificate to present to the server.
- **Load Balance:** Select a load-balancing algorithm using the Algorithm pull-down menu. This choice determines the method and prioritization for distributing connections or HTTP requests across available servers. The available options are:
    - **Consistent Hash:** New connections are distributed across the servers using a hash that is based on a key specified in the field that appears below the LB Algorithm field. This algorithm inherently combines load balancing and persistence, which minimizes the need to add a persistence method. This algorithm is best for load balancing large numbers of cache servers with dynamic content. It is ‘consistent’ because adding or removing a server does not cause a complete recalculation of the hash table. For the example of cache servers, it will not force all caches to have to re-cache all content. If a pool has nine servers, adding a tenth server will cause the pre-existing servers to send approximately 1/9 of their hits to the newly-added server based on the outcome of the hash. Hence persistence may still be valuable. The rest of the server’s connections will not be disrupted. The available hash keys are:
        - **Custom Header:** Specify the HTTP header to use in the Custom Header field, such as Referer. This field is case sensitive. If the field is blank or if the header does not exist, the connection or request is considered a miss, and will hash to a server.

- **Source IP Address** of the client.
- **Source IP Address and Port** of the client.
- **HTTP URI,** which includes the Host header and the Path. For instance, www.avinetworks.com/index.htm

- **Fastest Response:** New connections are sent to the server that is currently providing the fastest response to new connections or requests. This is measured as time to first byte. In the End to End Timing chart, this is reflected as Server RTT plus App Response time. This option is best when the pool's servers contain varying capabilities or they are processing short-lived connections. A server that is having issues, such as a lost connection to the data store containing images, will generally respond very quickly with HTTP 404 errors. It is best practice when using the Fastest Response algorithm to also enable the Passive Health Monitor, which recognizes and adjusts for scenarios like this by taking into account the quality of server response, not just speed of response.



- **Fewest Servers:** Instead of attempting to distribute all connections or requests across all servers, Avi Vantage will determine the fewest number of servers required to satisfy the current client load. Excess servers will no longer receive traffic and may be either de-provisioned or temporarily powered down. This algorithm monitors server capacity by adjusting the load and monitoring the server's corresponding changes in response latency. Connections are sent to the first server in the pool until it is deemed at capacity, with the next new connections sent to the next available server down the line. This algorithm is best for hosted environments where virtual machines incur a cost.
- **Least Connections:** New connections are sent to the server that currently has the least number of outstanding concurrent connections. This is the default algorithm when creating a new pool and is best for general-purpose servers and protocols. New servers with zero connections are introduced gracefully over a short period of time via the Connection Ramp setting in the Step 3: Advanced tab, which slowly brings the new server up to the connection levels of other servers within the pool.



- **Least Load:** New connections are sent to the server with the lightest load, regardless of the number of connections that server has. For example, if an HTTP request that will require a 200k response is sent to a server and a second request that will generate a 1k response is sent to a server, this algorithm will estimate that —based on previous requests— the server sending the 1k response is more available than the one still streaming the 200k of data. The idea is to ensure that a small and fast request does not get queued behind a very long request. This algorithm is HTTP specific. For non-HTTP traffic, the algorithm will default to the Least Connections algorithm.
- **Round Robin:** New connections are sent to the next eligible server in the pool in sequential order. This static algorithm is best for basic load testing, but is not ideal for production traffic because it does not take the varying speeds or periodic hiccups of individual servers into account.

There are several other factors beyond the Load Balancing algorithm that can affect connection distribution, such as Connection Multiplexing, server Ratio, Connection Ramp, and server Persistence.

- **Persistence:** By default, Avi Vantage will load balance clients to a new servers each time the client opens a new connection to a virtual sService, and there is no guarantee that the client will reconnect to the same server that they were previously connected to. A Persistence Profile ensures that subsequent connections from the same client will connect to the same server. Persistence can be thought of as the opposite of load balancing: a client's first connection to Avi Vantage is load balanced; thereafter, that client and any connections made by it will be persisted to the same server for the desired duration of time. Persistent connections are critical for most servers that maintain client session information locally. For example, many HTTP applications will keep a user's information in memory for 20 minutes, which allows the user to continue their session by reconnecting to the same server.

  As a best practice, HTTP virtual services requiring persistence should use HTTP Cookies, while general TCP or UDP applications requiring persistence will use Source IP. For more information on persistence types, see Persistence Profiles.

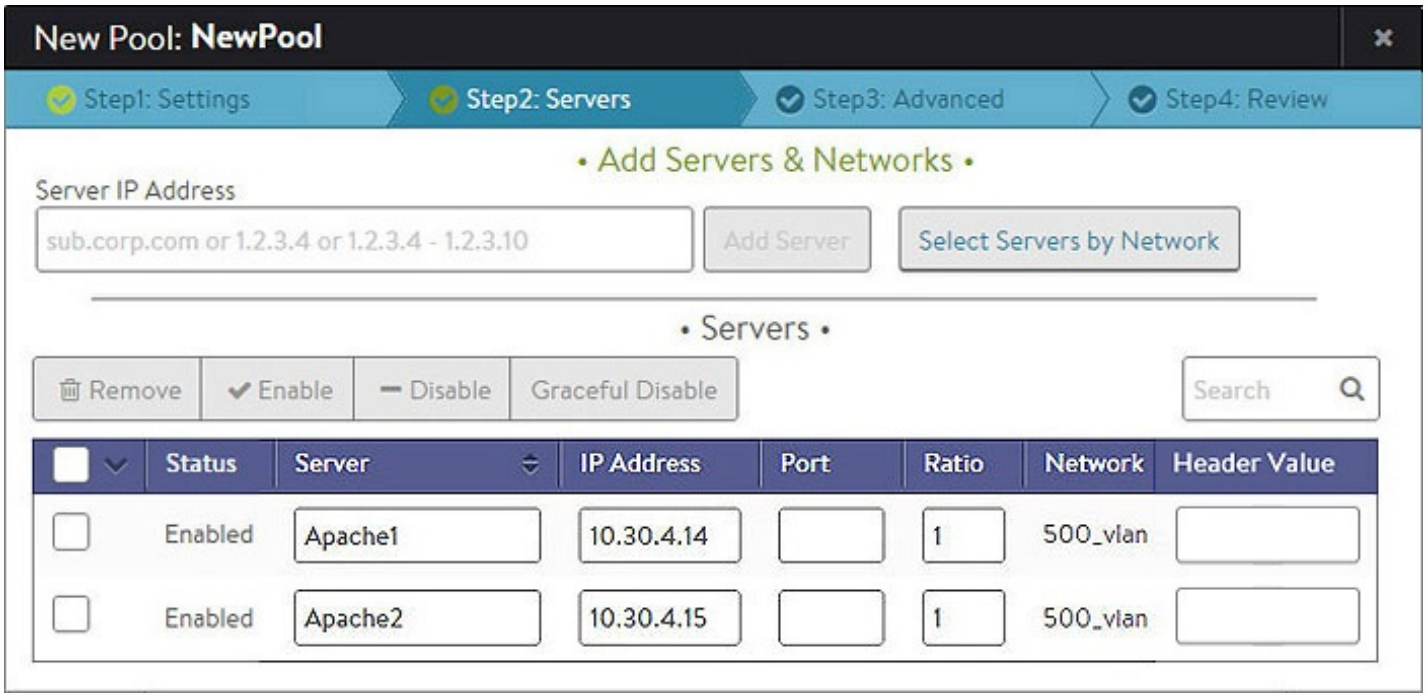
- **Health Monitor:** Vantage uses health monitors to generate synthetic connections or requests to servers to ensure the integrity of the server's health. You may add one or more health monitors to the pool by clicking the green add button and either selecting a health monitor or clicking the create health monitor button. You may also:
  - Disassociate a health monitor from the pool by clicking the trash can icon to the right of the monitor name.

- Edit an associated health monitor by clicking the blue edit pencil icon to the right of the associated monitor's name.

- **Passive Health Monitor:** A passive health monitor watches all client interactions with the site. If servers are sending errors (such 500 Busy or TCP connection errors), then the passive health monitor will reduce the amount of connections or requests sent to that server. The reduction percentage depends on the number of servers available within the pool. As the server responds satisfactorily to the throttled requests directed to it, the passive health monitor will restore the server to full traffic volume. You may use this monitor in conjunction with any other health monitors. Errors are defined in the Analytics profile assigned to the virtual service. Best practice is to ensure Passive Health Monitor is enabled in addition to any synthetic check that may also be configured.

Create Pool: 2 Servers

The Servers tab contains the server list for the pool.



Add Servers

- **IP Address, Range, or DNS Name:** Add one or more servers to the pool using one or more of the listed methods. The example below shows servers created using multiple methods.

Displaying 3 item(s)

	Status	Server	IP Address	Port	Ratio	Network	Header Value
☐	Enabled	10.1.1.100	10.1.1.100		1		
☐	Enabled	server2	10.10.5.76		1		
☐	Enabled	Apache1	10.30.4.14		1	500_vlan	

- **Add by IP Address:** Enter the IP address for the server that you want to include in the Address field, then click the green Add Server button. You may also enter a range of IP addresses via a dash, such as 10.0.0.1-10.0.0.20.
- **Add by DNS Resolvable Name:** Enter the name of the server in the Address field. If the server successfully resolves, the IP address will appear and the Add Server button will change to green. Click Add Server to include in the pool server list. See Add Servers by DNS
- **Select Servers by Network:** This option is only available if Vantage has read or write access to the cloud orchestrator. Click the Select Servers by Network button to open a list of reachable networks. Select a network to open a list of servers (virtual machines) available on that network. Filter the search for servers, such as searching for “apache” then select all matching servers. Click the green Add Servers button to include the new servers in the pool.Adding servers using the Select Servers by Network method allows Vantage to provide significantly richer information regarding the server. Avi Vantage is able to query the virtualization orchestrator for the virtual machine’s CPU, memory, and disk utilization. This is useful for better load balancing and visibility, and is the best-practice method. Adding servers by IP address or name will not provide this information. After a server has been added via the method, the server’s Network column in the server list table will be populated with the network or port group. See Select Servers by Network for more help.

- **IP Group:** Rather than add servers to an individual pool, server IP addresses may be added to an IP Group. This may be useful if the same group is used elsewhere for IP whitelists, DataScripts, or similar automation purposes. Many common pool features are unavailable when using this method, such as manually disabling a server, setting a specific service port, or setting a ratio. The IP Group method for adding servers may not be used with other methods.

Servers

- **Changing Server Status:** Adding servers to the pool populates the primary table of the

Servers tab, where you may now remove, enable, disable, or gracefully disable them. Changes to server status take effect immediately when you save your changes.

- **Remove:** Select one or more servers to remove from the pool. This will immediately reset any existing client connections for these servers and purge the server from the pool's list.
- **Enable:** Select one or more disabled servers, and then reactive them by clicking the Enable button. Enabling a server makes that server immediately available for load balancing, provided it passes its first health check.
- **Disable:** Select one or more enabled servers to disable. Avi Vantage immediately marks a disabled server as unavailable for new connections and resets any existing client connections. A server will not receive health checks while it is in a Disabled state.
- **Graceful Disable:** Similar to the Disable option, this also puts a server in an unavailable mode in that it will no long receive new connections; however, existing connections will be allowed to continue for the specified duration of time, in minutes. During this time, clients can finish their connections or data transfers. Any remaining open connections are reset when the timer expires. Valid timeouts range from 0 (immediate disable) to 60 minutes. A server will not receive health checks while it is in a gracefully-disabled state.

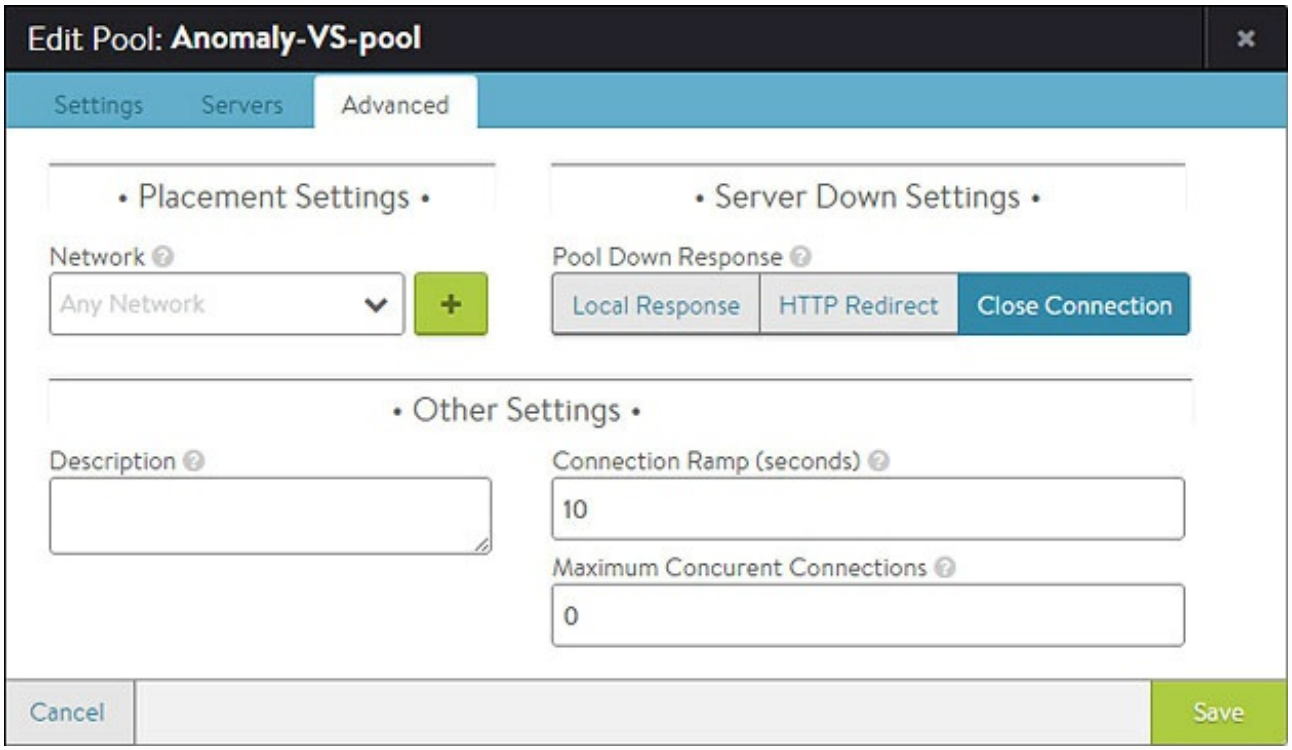
- **Editing Servers:** Servers added to the pool can be modified by editing their IP Address, Port, or Ratio fields.
  - **Status:** A server may be in an Enabled or Disabled.
  - **Server:** Name of the server (or the IP address, if the server was added manually).
  - **IP Address:** Changing the IP address for an existing server will reset any existing connections for the server.
  - **Port:** This optional field overrides the default service port number for the pool by giving server a specific port number that might differ from the other servers in the pool.
  - **Ratio:** This optional field creates an unequal distribution of traffic to a server relative to its peers. Ratio is used in conjunction with the Load Balancing algorithm. For example, If Server A has a Ratio of two and Server B has a Ratio of one, then Server A will receive two connections for every one connection that is sent to Server B. The Ratio may be any number between 1 and 20.

transparent
servers

- **Network:** Shows networks of the servers in the pool if Select Servers by Network was used.
- **Header Value:** This special field is used by the Custom HTTP Header persistence. Each server may be statically allocated an identifier, such as s1, s2, etc. If the selected client header exists, and the header value is s1, this server will receive the connection or request.

Create Pool: 3 Advanced

The Advanced tab of the Pool Create/Edit popup specifies optional settings for the pool.



Placement Settings

- **Server Network:** In some scenarios, a server may exist in multiple networks. Similarly, a network may have multiple IP subnets or a single subnet may exist in multiple networks. For example, VMware servers may have multiple Port Groups assigned to a single subnet, or a single Port Group is assigned to multiple subnets. Normally, Vantage will try to determine the network for the servers. However in scenarios where it cannot determine which network to use, an administrator may be required to manually select the server network to use.

Server Down Settings

- **Pool Down Action:** If all servers in a pool are down, the default behavior of the virtual

service is to close new client connection attempts by issuing TCP resets or dropping UDP packets. Existing connections are not terminated, even though their server is marked down. The assumption is the server may be slow but may still be able to continue processing the existing client connection.

- HTTP Local Response: returns a simple web page. Specify a status code of 200 or 503. If a custom HTML file has not been uploaded to Vantage, Vantage will return a basic page with the error code.
- HTTP Redirect: returns a redirect HTTP response code, including a specified URL.
- Close Connection: the default behavior of a pool for new client connections when all servers are down.
- Backup Pool: sends new connections to the specified pool. If servers within the original pool come online, connections to the backup pool will remain on that pool for their duration.

Other Settings

- **Disable Port Translation:** This feature is for virtual services that are listening on multiple service ports, such as Microsoft Lync, which has multiple listener ports. Instead of having all connections directed to a single port on the server (defined by the pool's Default Server Port or the server's optional Port field), they will be sent to the same port that they were received on the virtual service.

- **Description:** Enter an optional description of up to 256 characters in this field. This field is for user convenience only.

- **Connection Ramp:** Enabling this option by entering a number larger than 0 allows a graceful increase in the number of new connections sent to a server over the specified time period. For example, assume that the load balancing algorithm is set to Least Connections and a pool has two servers with 100 connections each. Adding a third server would immediately overwhelm that third server by immediately sending the next 100 consecutive connections to it. Setting a Connection Ramp adds traffic to a new server in a manner similar to using a Ratio. Over the specified period of time, the new server will receive an ever-increasing ratio of traffic in relation to its peers. For instance, setting the ramp to 4 seconds means that the new server will receive 1/4 of the traffic it would normally be given for the 1st second. By the 2nd second, the server will be receiving 1/2 the traffic it might otherwise be given. After the 4-second ramp time has elapsed, the server will receive the normal amount of traffic as determined by the load balancing algorithm.

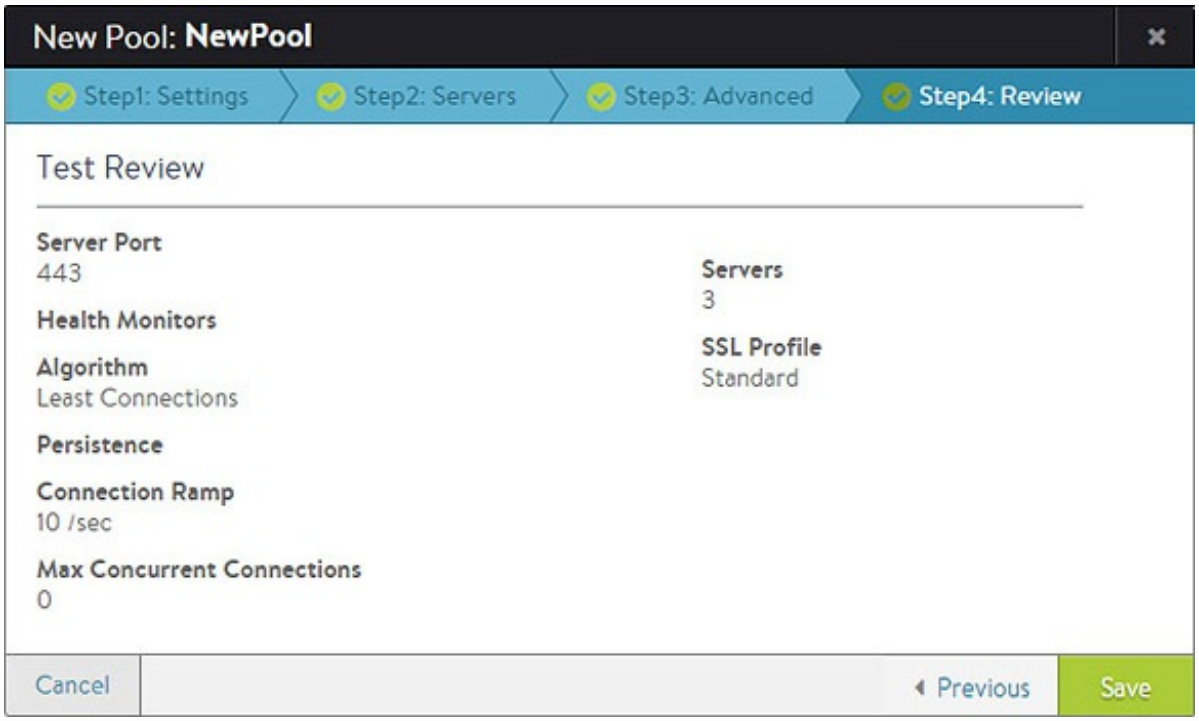
  Setting a Connection Ramp adds traffic to a new server in a manner similar to using a Ratio. Over the specified period of time, the new server will receive an ever-increasing ratio of traffic in relation to its peers. For instance, setting the ramp to 4 seconds means that the new

server will receive 1/4 of the traffic it would normally be given for the 1st second. By the 2nd second, the server will be receiving 1/2 the traffic it might otherwise be given. After the 4-second ramp time has elapsed, the server will receive the normal amount of traffic as determined by the load balancing algorithm.

- **Max Connections per Server:** Specify the maximum number of concurrent connections allowed for a server. If all servers in the pool reach this maximum the virtual service will send a reset for TCP connections or silently discard new UDP streams unless otherwise specified in the Pool Down Action, described above. As soon as an existing connection to the server is closed, that server is eligible to receive the next client connection. Valid values are 0, which disables the connection limit, or any number from 50 to 10,000.

Create Pool: 4 Review

The Review tab displays a summary of the information entered in the previous pool creation tabs.



Review this information and then click Save to finish creating the pool. If needed, you may return to any previous step by clicking the appropriate tab at the top of the popup window.

Servers

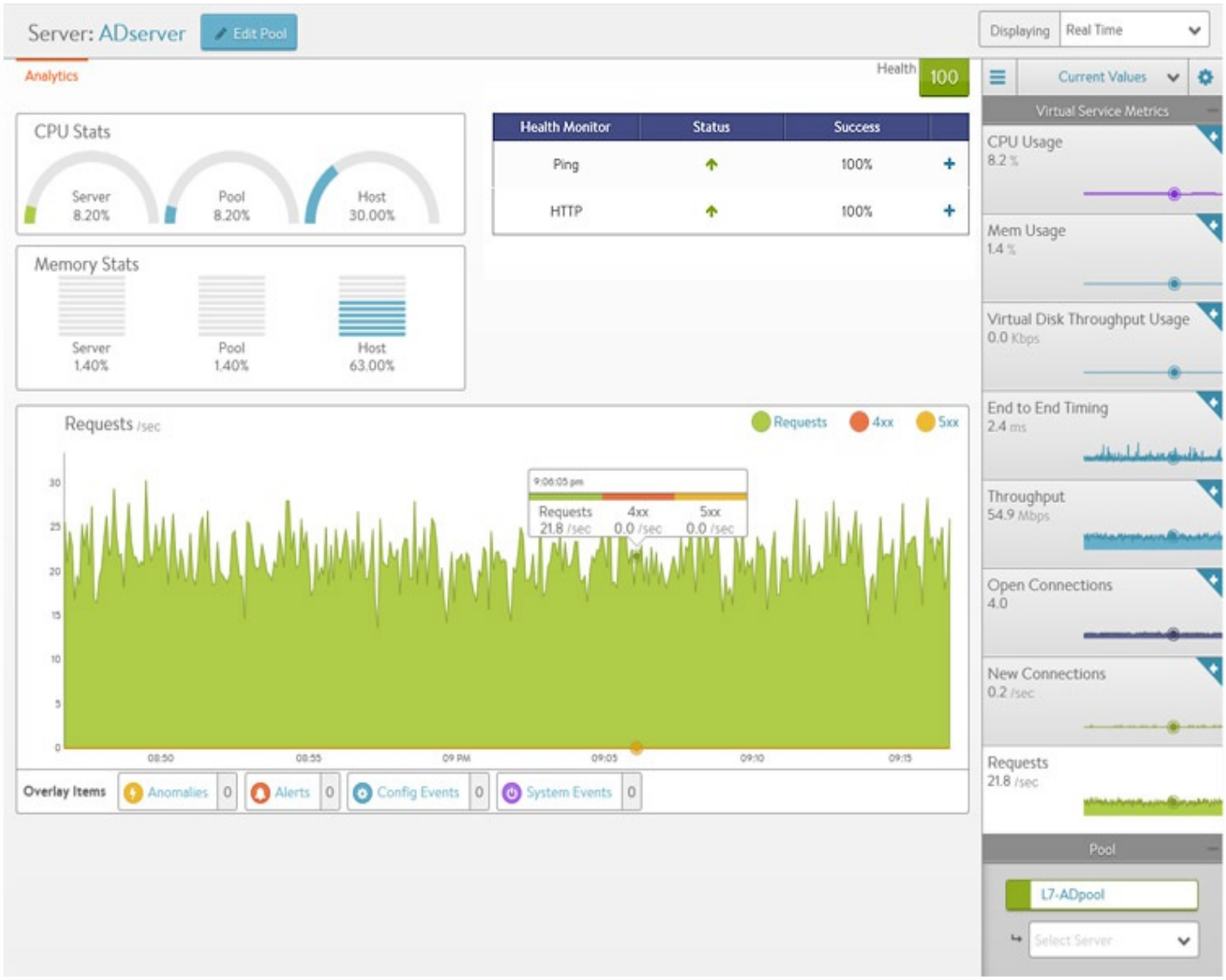
Information for each server within a pool is available within the Server Details Page. This page

allows views into correlation between server resources, application traffic, and response times.

Server Page

The Server Page may be access by clicking on the server's name from either the*Pool > Servers *page or the *Pool > Analytics' Servers* tile. When viewing the Server Details page, the server shown is within the context of the pool it was selected within. Rephrased, if the server (IP:Port) is a member of two or more pools, the stats and health monitors shown are only for the server within the context of the viewed pool.

Not all metrics within the Server Page are available in all environments. For instance, servers that are not virtualized or hooked into a hypervisor are not able to have their physical resources displayed.



The statistics can be changed or skewed by switching between Average Values, Peak Values, and Current Values. To see the highest CPU usage over the past day, change the time to 24 hour and the Value to Peak. This will show the highest stats recorded during the past day.

- **CPU Stats**: The CPU Stats box shows the CPU usage for this Server, the average during this time period across all servers in the pool, and the hypervisor host.
- **Memory Stats**: The memory Stats box shows the Memory usage for this server, the average during this time period across all servers in the pool, and the hypervisor host.
- **Health Monitor**: This table shows the name of any health monitors configured for the pool. The Status column shows the most current up or down health of the server. The Success column shows the percentage of health monitors that passed or failed during the display time frame. Clicking the plus will expand the table to show more info for a down server. See Why Is a Server Marked Down.
- **Main Panel**: The large panel shows the highlighted metric, similar to the VS Details and Pool Details pages. Overlay Items shows Anomalies, Alerts, Config Events, and System Events that are related to this server within the pool.

- **Pool Tile Bar**: The Pool in the top right bar shows the health of the pool. This can also be used to jump back up to the Pool Page. Under the pool name is a pull-down menu that allows quick access to jumping to the other servers within the Pool.
- **Metrics Tile Bar**: The metrics options will vary depending on the hypervisor Avi is plugged into. For non-virtualized servers, the metrics are limited to non-resource metrics, such as End to End Timing, Throughput, Open Connections, New Connections, and Requests. Other metrics that may be shown include CPU, Memory and Virtual Disk Throughput.

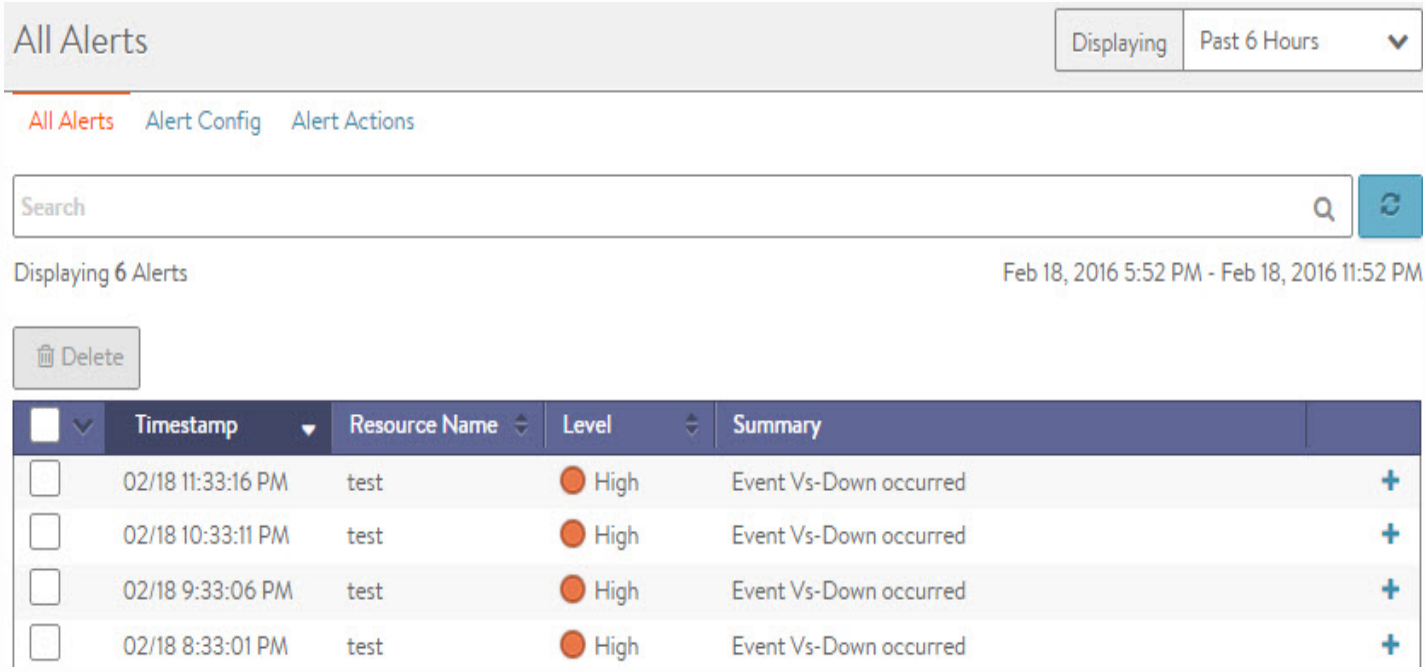
Alerts

Events are generated automatically within by Vantage to create a record of notable activities throughout the system. These may be Vantage errors, client traffic deviations, or many others.

Alerts may be generated when a trigger such as a system event occurs. If all conditions within the Alert Config are met, an alert is created. It may be used to actively or passively notify administrators, and may be deleted after being viewed or after an elapsed period of time.

All Alerts

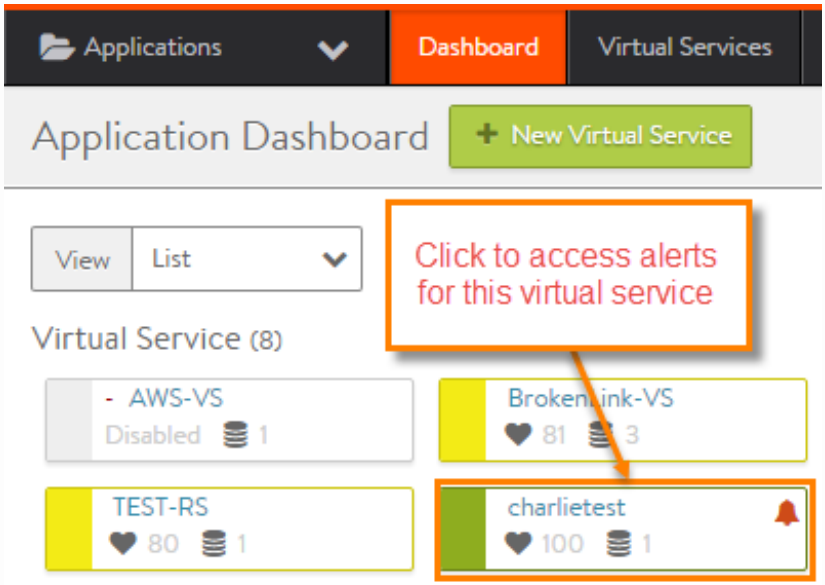
The Operations > Alerts > All Alerts page shows all Alerts across Avi Vantage.



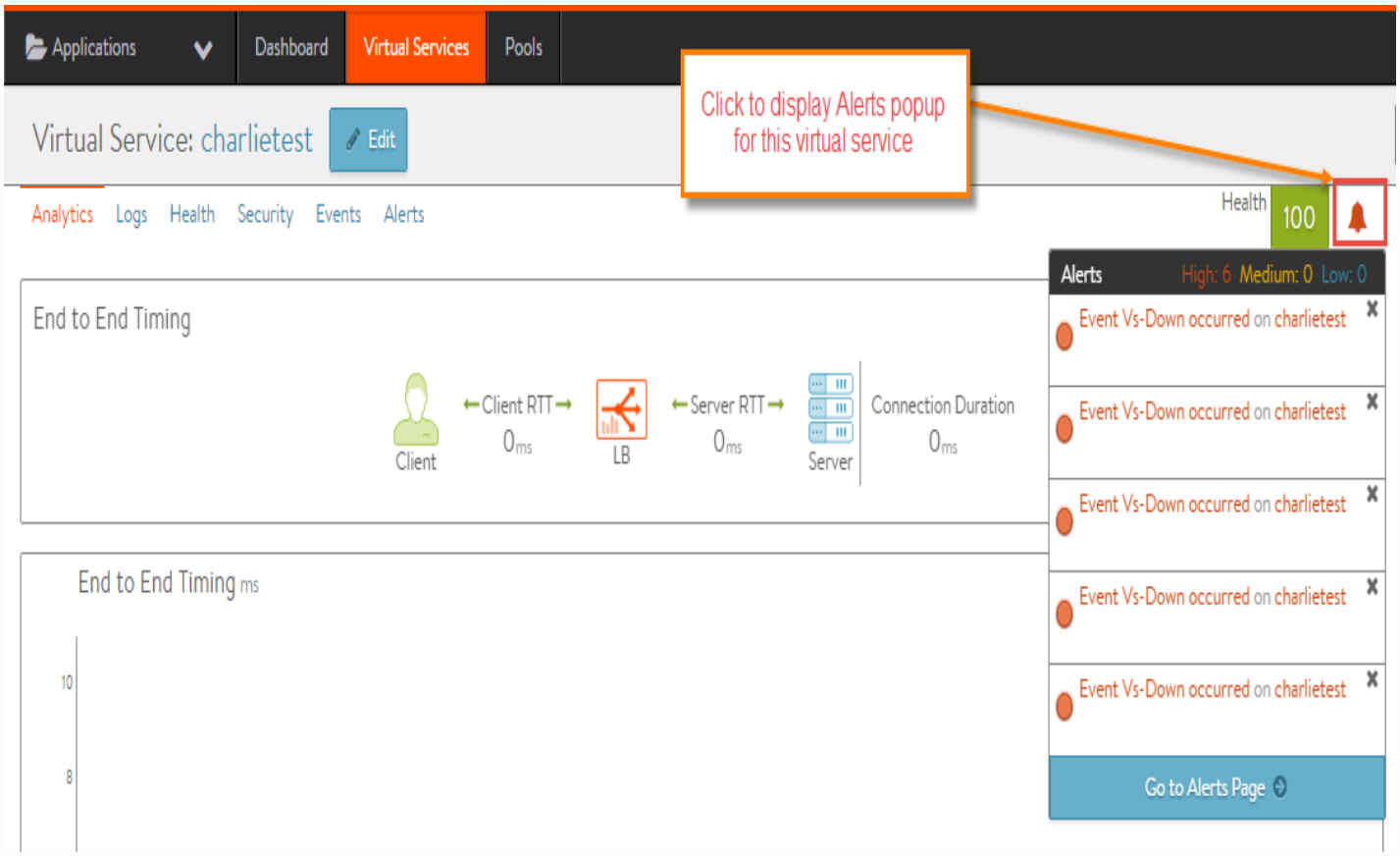
The functionality on this page is identical to that of the alerts tab found in the virtual services and pools pages; however, those pages are filtered to only show alerts related to those objects. See the Virtual Service Alerts tab for more information on navigating the Alert options.

Alerts for Specific Virtual Service

If the virtual service has alerts, an alerts icon appears on the virtual service icon on the Dashboard. Click the virtual service icon to access details for the virtual service, including its alerts:



Click Alerts to display the alerts log for the virtual service, or click the alerts icon next to the health score to display a popup list of the mot recent alerts.



Alert Configuration

Alerts filter events to highlight important information. Avi Vantage includes a number of default

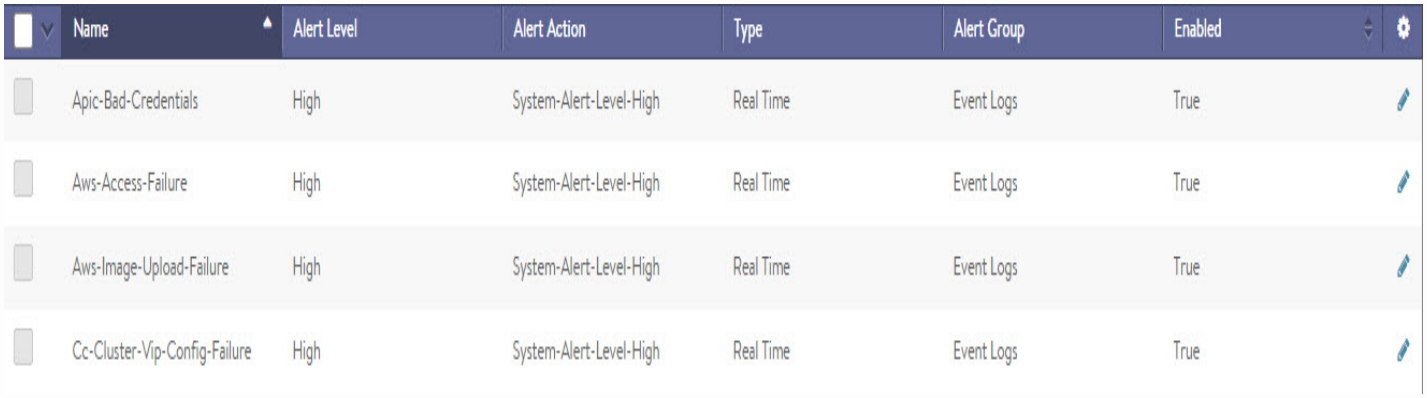
alerts via the Alert Config page. The default alerts may not be deleted, but can be modified or disable as needed. Alert Configs determine whether or not an alert should be generated.

Alert Config Page

The Alert Config page has the following options:

- **Search:** Search against the alert names in the table.
- **Create:** Opens the New Alert Config popup.
- **Edit:** Opens the Edit Alert Config popup.
- **Delete:** Remove the selected Alert configuration.

The table on this page contains the following information for each Alert:

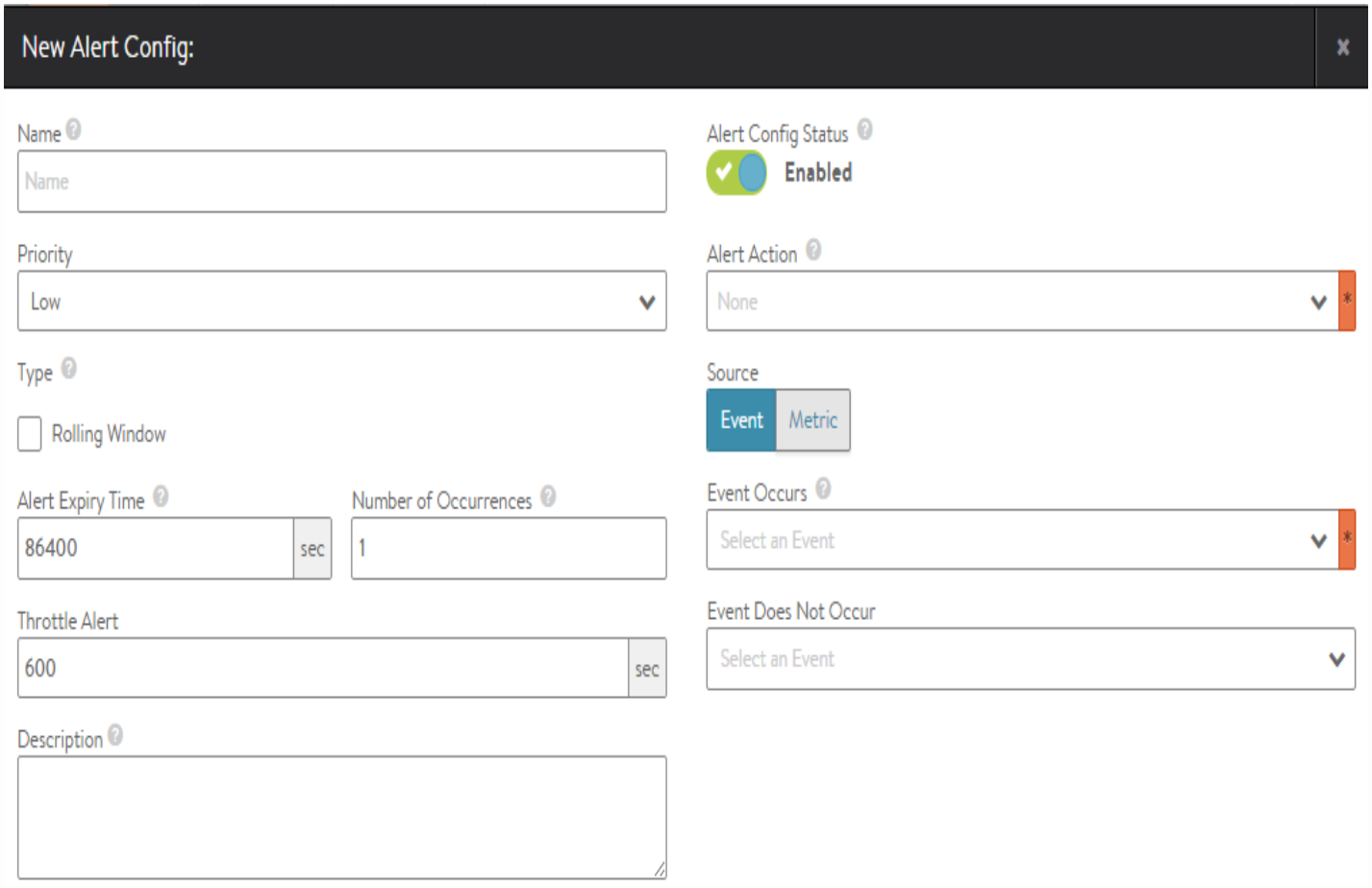


Name	Alert Level	Alert Action	Type	Alert Group	Enabled
Apic-Bad-Credentials	High	System-Alert-Level-High	Real Time	Event Logs	True
Aws-Access-Failure	High	System-Alert-Level-High	Real Time	Event Logs	True
Aws-Image-Upload-Failure	High	System-Alert-Level-High	Real Time	Event Logs	True
Cc-Cluster-Vip-Config-Failure	High	System-Alert-Level-High	Real Time	Event Logs	True

- **Name:** Name of the Alert.
- **Alert Level:** Level, or severity, of the Alert (Low, Medium, or High).
- **Alert Action:** Specifies the alert action that is triggered when the Alert is generated.
- **Type:** Whether the Alert is Real Time (5-second window) or Rolling Window (5-minute window).
- **Alert Group:** Whether the Alert applies to applications (App Log), client connections (Conn Logs), or Events (Event Logs)
- **Enabled:** Whether the Alert is enabled or disabled.

Create Alert Configuration

The New and Edit Alert Config popups share the same interface.



To create or edit an Alert notification:

- **Name:** Enter a unique name for the Alert Config.
- **Priority:** Every alert is classified with a priority level, which is used with alert actions. For example, all alerts marked as high could trigger an email notification. Alerts in the Vantage UI will indicate the level of severity. The following levels are available:
  - **External:** This alert will trigger an alert action, but will not create a local alert on Vantage.
  - **High:** Indicates a critical problem.
  - **Medium:** Indicates a less-severe problem.
  - **Low:** Indicates a minor problem.
- **Event:** If the selected event occurs, generate an alert.
- **Type:** Create an alert every time an event occurs, or only create an alert if an event occurs a specified number of times within a time window.
  - **Rolling Window Disabled:** When the event occurs, an alert is generated.
  - **Rolling Window Enabled:** An alert is generated if the Number of events is met or exceeded within the desired Time Window.
    - **Time Window:** Select a custom time frame. The default is 5 minutes.
- **Alert Expiry Time:** Alerts are intended to notify administrators of pressing issues that warrant attention, and many alerts are only actionable while the problem persists. Alerts are therefore automatically deleted after a period of time, with the underlying events and logs

serving as the definitive long-term records. An alert may be manually dismissed using the Alerts tab for the selected item (such as a virtual service).

- **Number of Occurrences:** During the Rolling Window timeframe, the event must occur this many times to trigger the alert.
- **Throttle Alert**: Sets the duration of time within which an alert may only be created once.
- **Description:** Enter a custom description for the alert.
- **Alert Config Status**: Disable an Alert Config from generating new alerts.
- **Alert Action:** Points to the alert action group for the alert. The action object is configured to perform an action such as email, Syslog, etc.
- **Source**: Set the event or metric that will trigger the alert.
    - **Event**:
        - **Event Occurs**: If an event happens, such as login failure. This field is required when alerting on events.
        - **Event Does Not Occur**: If the event does not occur, trigger the event. This field is optional. When used, both this field and the Event Occurs must be true for the alert to trigger. This option is useful if expecting to see an event within a period of time defined via the Alert Threshold Duration. For instance, an Event Occurs is set to server down while Event Does Not Occur is set to server up. If the server bounces down and back up again within a 5 minute window, don't create an alert.
    - **Metric**:
        - **Object**: Inspect the metrics of an object such as virtual service, Service Engine, or pool.
        - **Instance**: Select the specific object of the type specified in the Object field.
        - **Metric Occurs**: Select a metric which is used as the trigger for the alert.
        - **Comparator**: The value returned by the Metric Occurs must match the greater than, less than, or equal to the Value field.
        - **Value**: The numeric value the metric is compared against. Some metrics will be percentage from 0 to 100, while throughput is in bits per second, such as 1000000.
        - **Duration**: The length of time during which the metric must be true. Generally this means an average of the metric over this period of time. Keep in mind some metrics are not collected at an aggressive rate. For instance, vCenter updates server CPU information every five minutes. Therefore duration should be longer than 600 (seconds).

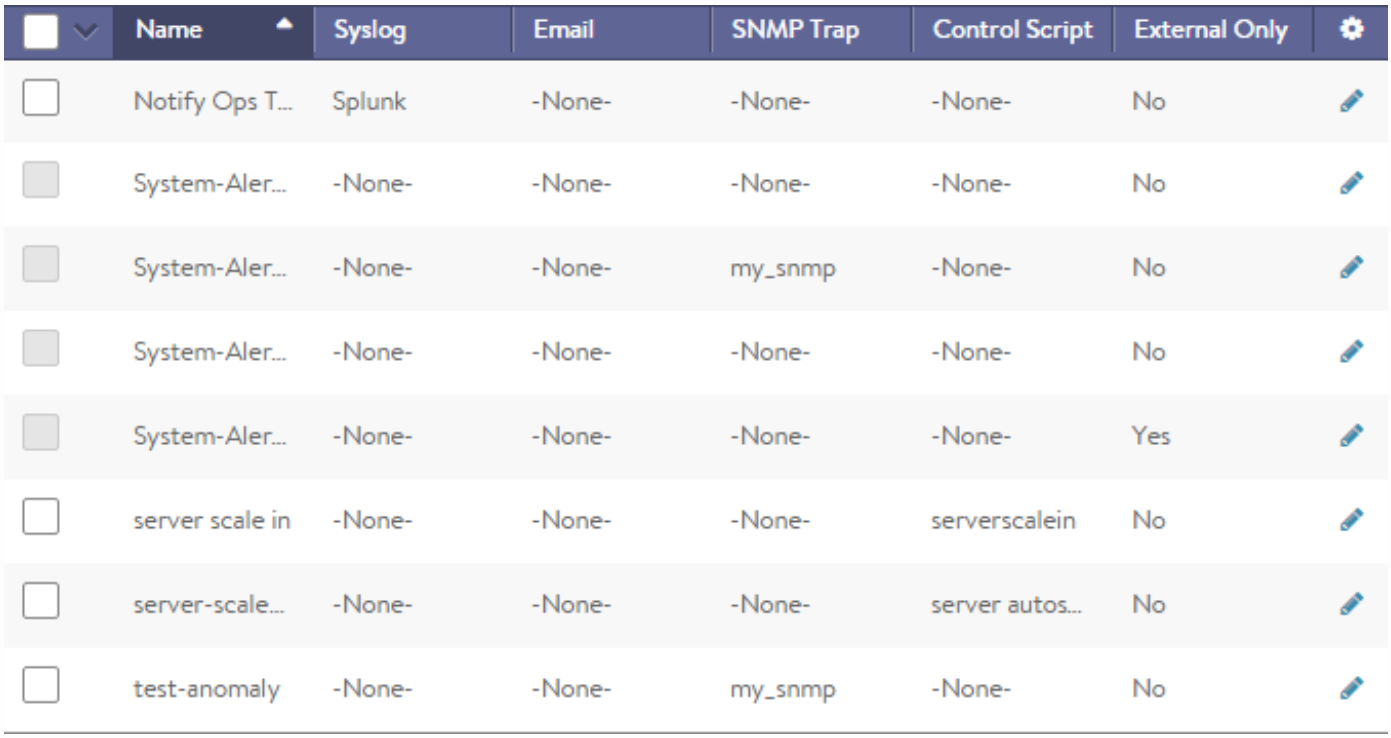
Alert Actions

When an alert is generated within Vantage, it is immediately able to be viewed by an administrator. This means the administrator must sign into Vantage, navigate to the appropriate object and view the alerts to know if there is a serious issue. Alternatively, alert actions may be set up, which allow alerts to be proactively sent from Vantage to an external destination.

Alert Action Page

Select Operations > Alerts > Alert Actions to display the configured alert actions, which correspond to the three system alert levels (high, medium, and low). Alert actions can be set up to export alerts matching any of these three severity levels. The actions for these levels can be changed, but the levels themselves cannot be modified as they map to alert severity displayed throughout the Vantage web interface. This tab includes the following functions:

- **Search:** Search for the name of the alert action.
- **Create:** Opens the Create/Edit Alert Action popup.
- **Edit:** Opens the Create/Edit Alert Action popup.
- **Delete:** Remove the selected alert action.



Name	Syslog	Email	SNMP Trap	Control Script	External Only
Notify Ops T...	Splunk	-None-	-None-	-None-	No
System-Aler...	-None-	-None-	-None-	-None-	No
System-Aler...	-None-	-None-	my_snmp	-None-	No
System-Aler...	-None-	-None-	-None-	-None-	No
System-Aler...	-None-	-None-	-None-	-None-	Yes
server scale in	-None-	-None-	-None-	serverscalein	No
server-scale...	-None-	-None-	-None-	server autos...	No
test-anomaly	-None-	-None-	my_snmp	-None-	No

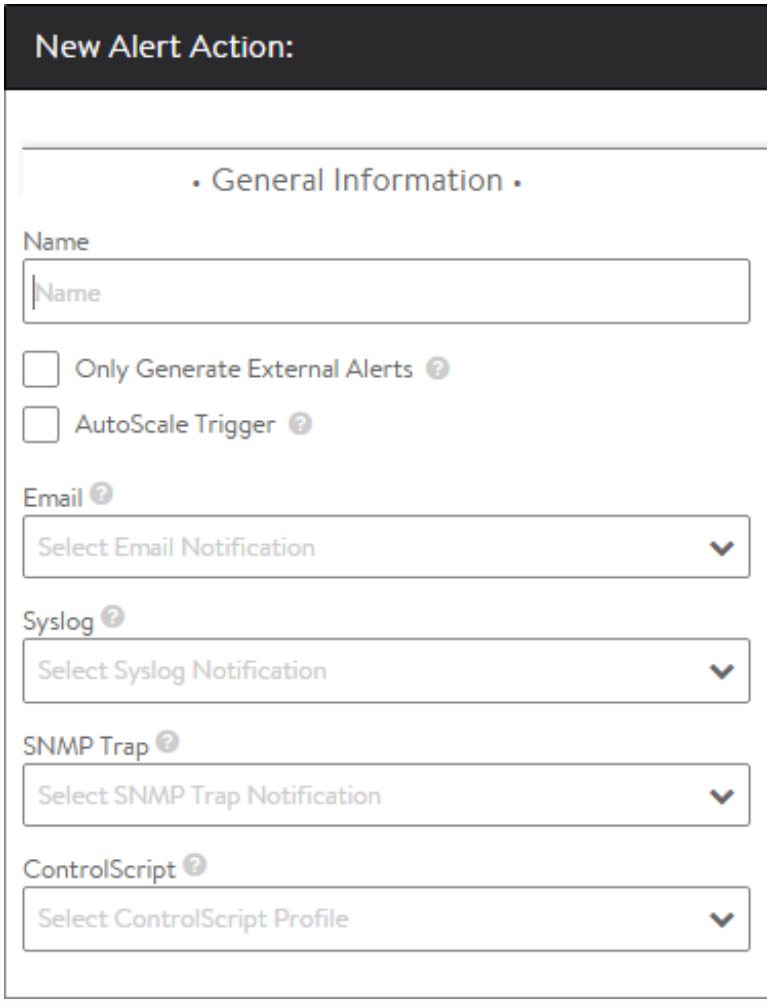
This example shows the system default alert actions (*System-Alert-*) and some custom alert actions configured by a Vantage user. The default system alert actions may be modified, but not deleted.

The table on this tab contains the following information for each alert action:

- **Name:** The name of the alert action.
- **Syslog:** Shows whether the alert action is configured to send Syslog messages.
- **Email:** Shows whether the alert action is configured to send Email messages.
- **SNMP Trap:** Shows whether the alert action is configured to send SNMP traps.
- **Control Script:** Shows whether the alert action is configured to trigger execution of a ControlScript.
- **External Only:** Shows whether the alert action is configured to execute an external script.

Create/Edit an Alert Action

Alert actions may be stored locally on Vantage, and they may also be sent to remote systems. To direct alerts to a remote system, first configure a Notification or a ControlScript which may be referenced by the alert action.



Alert action options that can be configured here:

- **Name:** User friendly name for the action alert.
- **Only Generate External Alert:** By default, Vantage generates a local alert visible to admins of Vantage, in addition to any external alert systems configured. When this option is enabled, alerts mapped to this alert action will not generate a local alert within Vantage. Alerts may still be sent externally via any of the methods below.
- **AutoScale Trigger:** Use the alert to trigger AutoScaling.
- **Email:** Select an existing Email Notification.
- **Syslog:** Select an existing Syslog Notification.
- **SNMP Trap**: Select an existing SNMP Trap Notification.
- **ControlScript**: Launch a custom script. ControlScripts are Python-based scripts executed on the Avi Controller. These scripts may make configuration changes to Vantage, or send data externally to a remote system. For instance, a ControlScript could instruct an orchestrator to

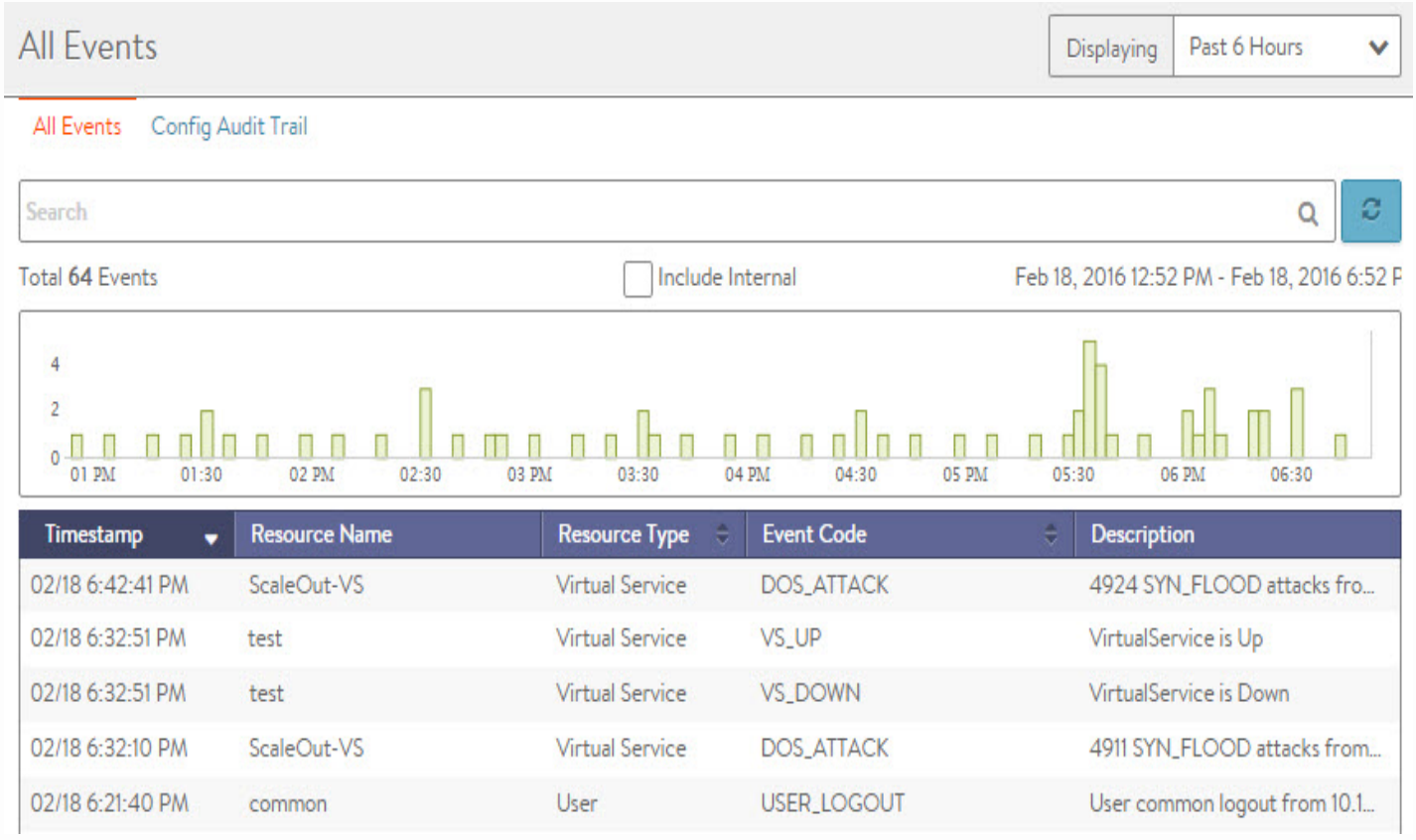
spin up new servers if all the servers are down within a pool or notify a security team if a virtual service is under a severe denial of service attack.

Events

Events are used throughout the system to provide a history of relevant changes that have occurred. Events are a permanent record, and can be used to generate Alerts which can take action on the Event. Events are viewable in the context of specific objects, such as a Virtual Service, a Pool, or a Server. Viewing Events from the Operations menu provides an unfiltered view of all Events across the system or the tenant.

All Events

Select Operations > Events to open the All Events page, which shows all Events for the entire Avi Vantage system. Events pages elsewhere within Avi Vantage will be filtered by the context of the section they exist within, such as only showing Events relevant to the Virtual Service they are being viewed within.



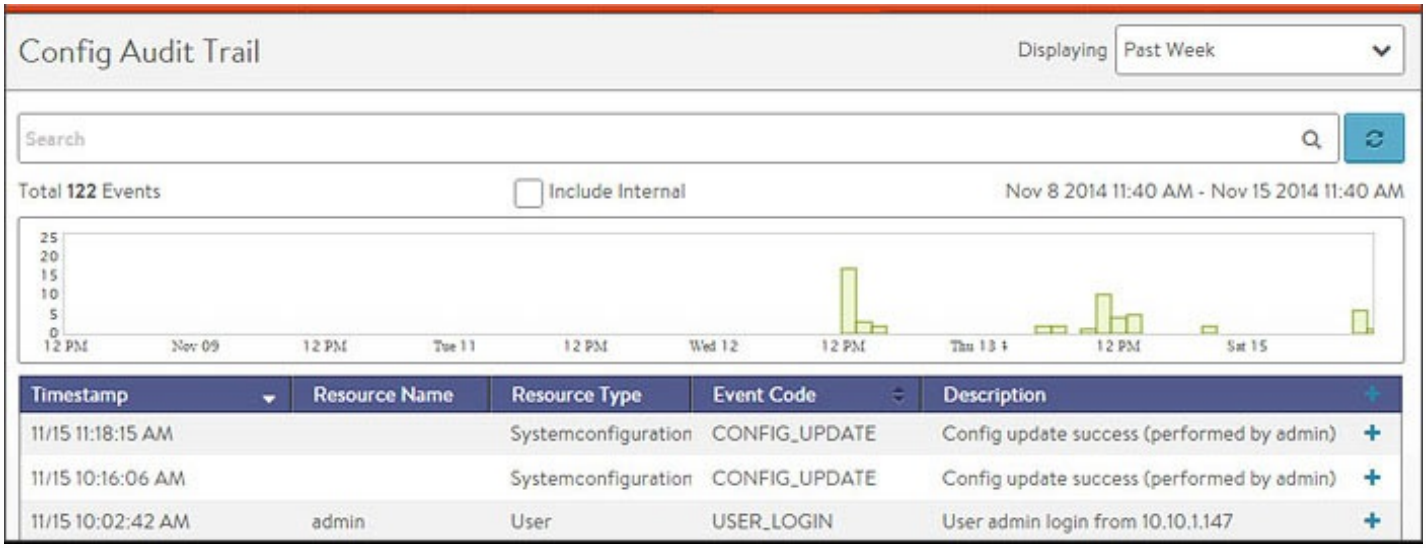
The top of this page displays the following items:

- **Search:** The Search field allows you to filter the audit trail using whole words contained within the individual events.

- **Refresh:** Updates the audit trail events that are displayed for the selected time frame.
- **Number:** Total number of entries being displayed. The date/time range of those events appear beneath the Search field on the left.
- **Clear Selected:** If filters have been added to the Search field, the Clear Selected X icon on the right side of the search bar will remove those filters. Each active search filter will also contain an X that can be clicked to remove a specific filter.
- **Histogram:** The Histogram shows the number of events over the period of time selected. The X-axis is time, while the Y-axis is the number of events during that bar's period of time.
    - Hovering the cursor over a Histogram bar displays the number of entries represented by that bar during a period of time.
    - Click and drag inside the histogram to refine the date/time period which further filters the events shown. When drilling in on the time in the Histogram, a Zoom to Selected link appears above the Histogram. This expands the drilled in time to expand to the width of the Histogram, and also changes the Displaying time pull-down menu to Custom. To return to the previously selected time period, use the Display menu to select a new time period.
- **Include Internal**: By default, a number of events are not shown as they tend to be noisy and less relevant for general purpose. Avi support may ask to enable this option when working with support to troubleshoot more esoteric issues.

Config Audit Trail

Select Operations > Events > Config Audit to open the Config Audit Trail page, which presents an audit trail of user activity events as well as changes to the system configuration over the period of time selected. This is a filtered view of the All Events, showing only events related to configuration changes.



The table at the bottom of the Config Audit Trail tab displays the events that matched the current time window and any potential filters. The following information appears for each event:

- **Timestamp:** Date and time the event occurred. Highlighting a section of the Histogram allows further filtering of events within a smaller time window.
- **Event Type:** This page is scoped to only show Configuration event types. Configuration events track changes to the Avi Vantage configuration. These changes may be made by an administrator (through the CLI, API, or GUI), or by automated policies.
- **Resource Name:** Name of the object related to the event, such as the Pool, Virtual Service, Service Engine, or Controller. For User_Login Events, this will show the username that attempted to log in.
- **Event Code:** A short event definition, such as User_Login, Config_Create, Config_Modify, or Config_Delete.
- **Description:** A complete event definition. For configuration events, the description will also show the username of the Avi Vantage account that made the change.
- **Expand/Contract:** Clicking the + plus or - minus sign for an event log either expands that event log to display more detail or contracts that event log to display only summary information. Clicking the plus and minus icons in the table header expands and collapses all entries in this tab.

For configuration events, expanding the event displays a difference comparison between the previous and new configurations.

- New fields will appear highlighted in green in the new configuration.
- Removed fields will appear highlighted in red.

- Changed fields will show highlighted in yellow.

Services

Alert actions may be configured to proactively send notifications to an administrator using the methods available on the Notifications page (email, syslog, SNMP).

Before creating an alert action, the notification destinations must be configured. These can be syslog servers, email servers and addresses, and SNMP trap servers. In addition to notifications, an alert action also can include running a ControlScript. Notifications need to be configured only if they are going to be used for an alert action.

Syslog Notifications

Alert actions may be configured to send alerts to one or more remote Syslog servers via TCP to the configured IP address and service ports. The alert action determines which log levels (high, medium, low) should be sent. Avi Vantage uses this process internally for receiving logs. This appears on the Syslog tab as System-Syslog.

Configuring Syslog notifications pushes alerts to Syslog servers. It does not export virtual service logs. These may be pulled from an external logging system via the API, or may be scripted to push from the Avi Controller.

Syslog Notifications Settings

Select Administration > Notifications > Syslog to open the Syslog Notifications tab. This tab includes the following functions:

- **Search:** Search through the configure list of Syslog entries.
- **Create:** Opens the Create Syslog Notification popup.
- **Edit:** Opens the Edit Syslog Notification popup.
- **Delete:** Remove the selected Syslog notifications. The default System-Syslog notification

may be modified, but not deleted.

The table on this tab contains the following information for each Syslog notification:

- **Name:** Name of the Syslog notification.
- **Syslog Server:** IP address or hostname of the destination Syslog server for the log entries. The server must be listening for TCP.
- **Port:** The service port number of the destination Syslog server.

Create Syslog Notification

The New Syslog and Edit Syslog Notification popups share the same interface.

To create or edit a system log notification:

- **Name:** Enter a unique name for the Syslog destination.
- **Syslog Server:** Enter either the IP address or hostname of the remote syslog server.
- **Port:** Enter the service port of the destination Syslog server. Avi Vantage uses TCP as the protocol for sending logs.

Email

Alert Actions may be configured to send alerts to administrators via email. These emails could be sent directly to administrators or to reporting systems that accept email. Either option requires the Vantage Controller to have a valid DNS and default gateway configured so it can resolve the destination and properly forward the messages. Information regarding the SMTP server and the sender must be configured in the Administration > Settings > Email/SMTP page.

Information regarding the SMTP server and the sender must be configured in the Administration >

Settings > Email/SMTP page.

Select Operations > Services > Email to open the Email tab. This tab includes the following functions:

- **Search:** Search through this list of email notification names.
- **Create:** Opens the Create/Edit Email Notification popup.
- **Edit:** Opens the Create/Edit Email Notification popup.
- **Delete:** Remove the selected email notifications.

The table on this tab contains the following information for each email notification:

- **Name:** Name of the email notification.
- **To Address:** The email address used as the primary destination for the Alert Action. Use comma separation for multiple email addresses.
- **CC Address:** The email address used as the secondary, or CC'd email destination for the Alert Action. Use comma separation for multiple email addresses.

SNMP Trap

Alerts may be sent via SNMP Traps using SNMP v2c.

SNMP Trap Settings

Select Operations > Services > SNMP Trap to open the SNMP tab. This tab includes the following functions:

- **Search:** Search through this list of SNMP Trap server names.
- **Create:** Opens the Create/Edit SNMP Trap popup.
- **Edit:** Opens the Create/Edit SNMP Trap popup.

- **Delete:** Remove the selected SNMP Trap server.

Creating an SNMP trap server presents the following options:

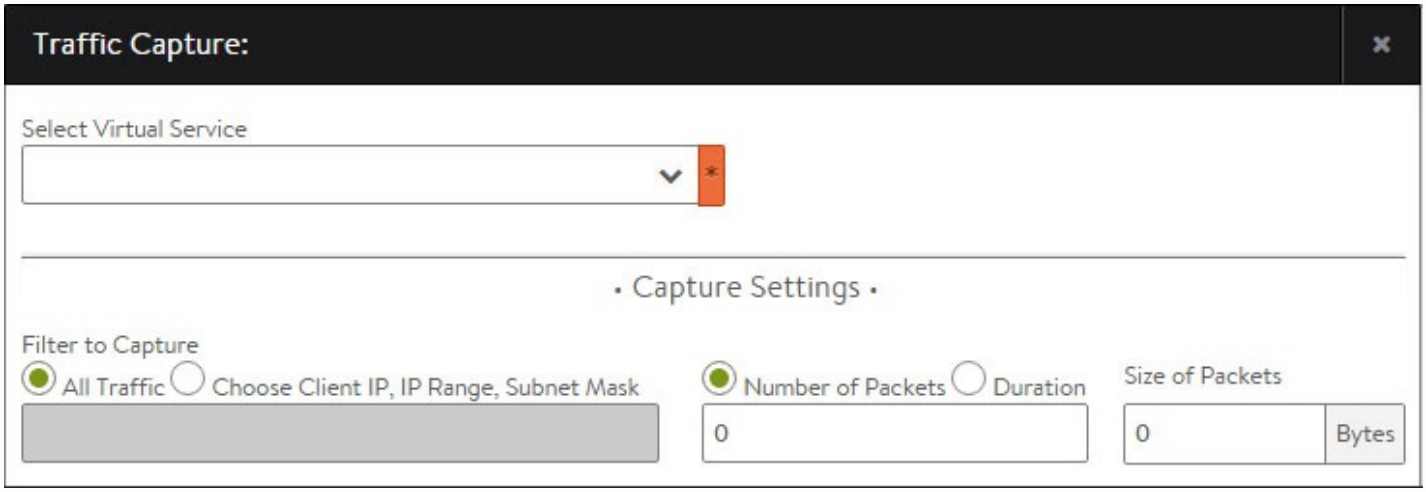
- **Name:** Name of the SNMP server.
- **Trap Server:** The IP address or hostname of the server.
- **SNMP Community:** Trap servers may require a community string, which provides a basic form of security for access to the server.

Traffic Capture

Most troubleshooting of connection or traffic data may be done quickly via Virtual Services logs. However, some troubleshooting may require full visibility into the packet transmission. Avi provides Packet Capture, which runs TCPdump against the desired Virtual Service. The packet capture is done on any Service Engines that may be hosting the VS, then collated in the completed capture. Packet capture may also be performed via the CLI.

Capture Configuration

The Capture Configuration section shows the parameters defined for any captures that are currently in progress. To begin a new capture, select the blue pencil icon on the right of the box.



- **Select Virtual Service:** The capture is executed against traffic for a virtual service. The capture includes both the client-to-SE and SE-to-server sides of the connection. It will automatically be performed on all SEs handling traffic for the virtual service.
- **All Traffic:** By default, all traffic is captured. Adding an optional filter will enable IP based filtering based on a single IP address 10.1.1.1, a space separated list, a range 10.1.1.1-10.1.1.255, or a subnet mask 10.1.1.1/24. This IP could be either the client or the server of

the connection.

- **Number of Packets / Duration:** It is highly recommended to set a limit to the duration of the capture. This limit may be either the max number of packets to receive, or the duration of time, in minutes. Once the limit has been reached, the capture will terminate and be sent to the Controller for processing and should be available shortly after.
- **Size of Packets:** Set the number of bytes to capture per packet. This is similar to the snaplen option in TCPdump.

When the capture is started, the Capture Configuration section indicates the progress of the capture. Once the capture is complete, it may still take a few minutes for the new capture to show up in the Completed Captures, as the Controller may need to collate data from multiple SEs and format the data, which is output as a pcap file.

Completed Captures

Once a traffic capture has completed, it will show in the Completed Captures table. This table shows the date, virtual service name and size. The far right column of the table contains an export icon, which allows administrators to download the pcap file. This file type can be viewed by common traffic capture utilities such as Wireshark.

Templates

A template is a named set of configuration items or option settings. Templates greatly simplify configuration by allowing the same template to be used by multiple objects. The settings in the template apply to all the objects that are associated with (bound to) the template.

A change to a template affects all objects that are associated with that template. A template does not need to be rebound to its objects after modification. The change is immediately applied to the objects that are associated with the template. For example, if the option to insert X-Forwarded-For headers is disabled in an HTTP application profile, that change is applied to all Virtual Services that use the profile.

Avi Vantage uses the following types of templates:

- Profiles ›
- Groups ›
- Security ›
- Scripts ›

Profiles

Avi Vantage relies on Templates and Profiles to group together collections of settings into a single, reusable object. A profile may be used by multiple objects such as Virtual Services referencing the same TCP Profile. By making a change to the TCP Profile, all associated Virtual Services will inherit the change.

Avi Vantage supports the following Templates and Profiles:

- Application Profile
- TCP/UDP Profile
- Persistence Profile
- Health Monitor Profile
- Analytics Profile

Application Profile

Application profiles determine the behavior of virtual services, based on application type.

The application profile types and their options are described in the following sections:

- HTTP Profile
- DNS Profile
- Layer 4 Profile
- Syslog Profile

Dependency on TCP/UDP Profile

The application profile associated with a virtual service may have a dependency on an underlying TCP/UDP profile. For example, an HTTP application profile may be used only if the TCP/UDP profile type used by the virtual service is set to type TCP Proxy. The application profile associated with a virtual service instructs the Service Engine (SE) to proxy the service's application protocol, such as HTTP, and to perform functionality appropriate for that protocol.

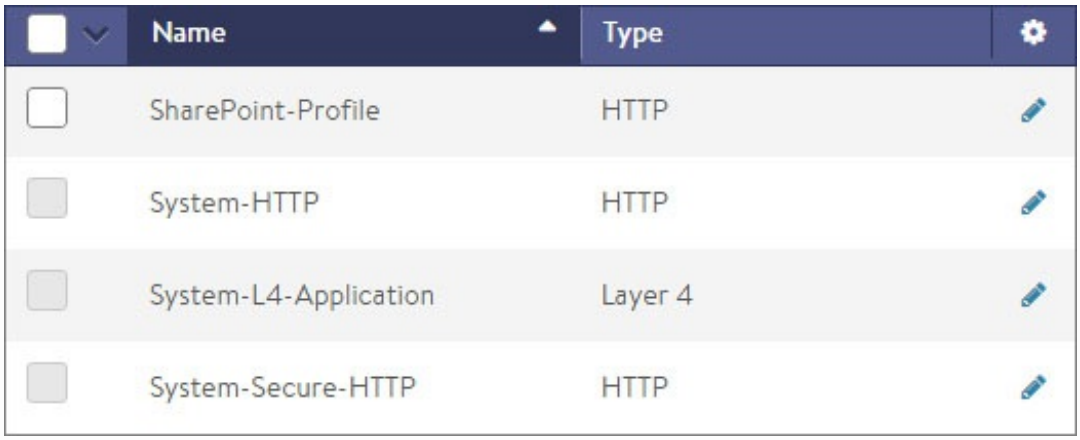
Application Profile Tab

Select Templates > Profiles > Applications to open the Application Profiles tab, which includes the following functions:

- **Search:** Search against the name of the profile.
- **Create:** Opens the Create Application Profile popup.
- **Edit:** Opens the Edit Application Profile popup.
- **Delete:** Removes an application profile if it is not currently assigned to a virtual service.

  Note: If the profile is still associated with any virtual services, the profile cannot be removed. In this case, an error message lists the virtual service that still is referencing the application profile.

The table on this tab provides the following information for each application profile:



- **Name:** Name of the Profile.
- **Type:** Type of application profile, which will be either:
  - **DNS:** Default for processing DNS traffic.
  - **HTTP:** Default for processing Layer 7 HTTP traffic.
  - **L4:** Catch-all for any virtual service that is not using an application-specific profile.
  - **Syslog:** Default for processing Syslog traffic.

Create/Edit an Application Profile

The Create Application Profile and Edit Application Profile screens share the same interface regardless of the application profile

chosen.

The initial settings for a new profile are similar regardless of the type of profile chosen:

- **Name:** Enter a unique name for the profile.
- **Description:** Enter an optional description for the profile.
- **Type:** Click the appropriate type button to select the application for this profile. Select L4 for none.

HTTP Profile

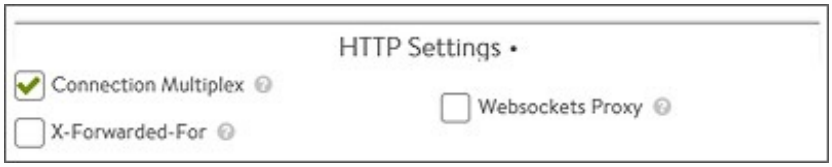
The HTTP application profile allows Vantage to be a proxy for any HTTP traffic. HTTP-specific functionality such as redirects, content switching, or rewriting server responses to client requests may be applied to a virtual service. The settings apply to all HTTP services that are associated with the HTTP profile. HTTP-specific policies or DataScripts also may be attached directly to a virtual service.

The HTTP profile contains these tabs:

- **General**
- **Security**
- **Acceleration**
- **DDoS**

HTTP General Tab

The general tab contains HTTP basic settings:



- **Connection Multiplex:** This option controls the behavior of HTTP 1.0 and 1.1 request switching and server TCP connection reuse. This allows Avi Vantage to reduce the number of open connections maintained by servers and better distribute requests across idle servers, thus reducing server overloading and improving performance for end-users. The exact reduction of connections to servers will depend on how long lived the client connections are, the HTTP version, and how frequently request/responses are utilizing the connection. It is important to understand that “connection” refers to a TCP connection, whereas “request” refers to an HTTP request and subsequent response. HTTP 1.0 and 1.1 allow only a single

request/response to go over an open TCP connection at a time. Many browsers attempt to mitigate this bottleneck by opening around six concurrent TCP connections to the destination web site. See Multiplex plus Persistence, below.

- **X-Forwarded-For:** With this option, Vantage will insert an X-Forwarded-For (XFF) header into the HTTP request headers when the request is passed to the server. The XFF header value contains the original client source IP address. Web servers can use this header for logging client interaction instead of using the layer 3 IP address, which will incorrectly reflect the Service Engine's source NAT address. When enabling this option, the XFF Alternate Name field appears, which allows the XFF header insertion to use a custom HTTP header name. If the XFF header or the custom name supplied already exists in the client request, all instances of that header will first be removed. To add the header without removing pre-existing instances of it, use an HTTP request policy.

- **WebSockets Proxy:** Enabling WebSockets allows the virtual service to accept a client's Upgrade header request. If the server is listening for WebSockets, the connection between the client and server will be upgraded. WebSocket is a full-duplex TCP protocol. The connection will initially start over HTTP, but once successfully upgraded, all HTTP parsing by Vantage will cease and the connection will be treated as a normal TCP connection.

- **Save:** Select another tab from the top menu to continue editing or Save to return to the Application Profiles tab.

Multiplex plus Persistence

Multiplexing behavior changes with server persistence enabled:

- **Multiplex enabled, Persistence disabled:** Client connections and their requests are decoupled from the server side of the Service Engine. Requests are load-balanced across the servers in the pool using either new or pre-existing connections to those servers. The connections to the servers may be shared by requests from any clients.

- **Multiplex enabled, Persistence enabled:** Client connections and their requests are sent to a single server. These requests may share connections with other clients who are persisted to the same server. Load balancing of HTTP requests is not performed.

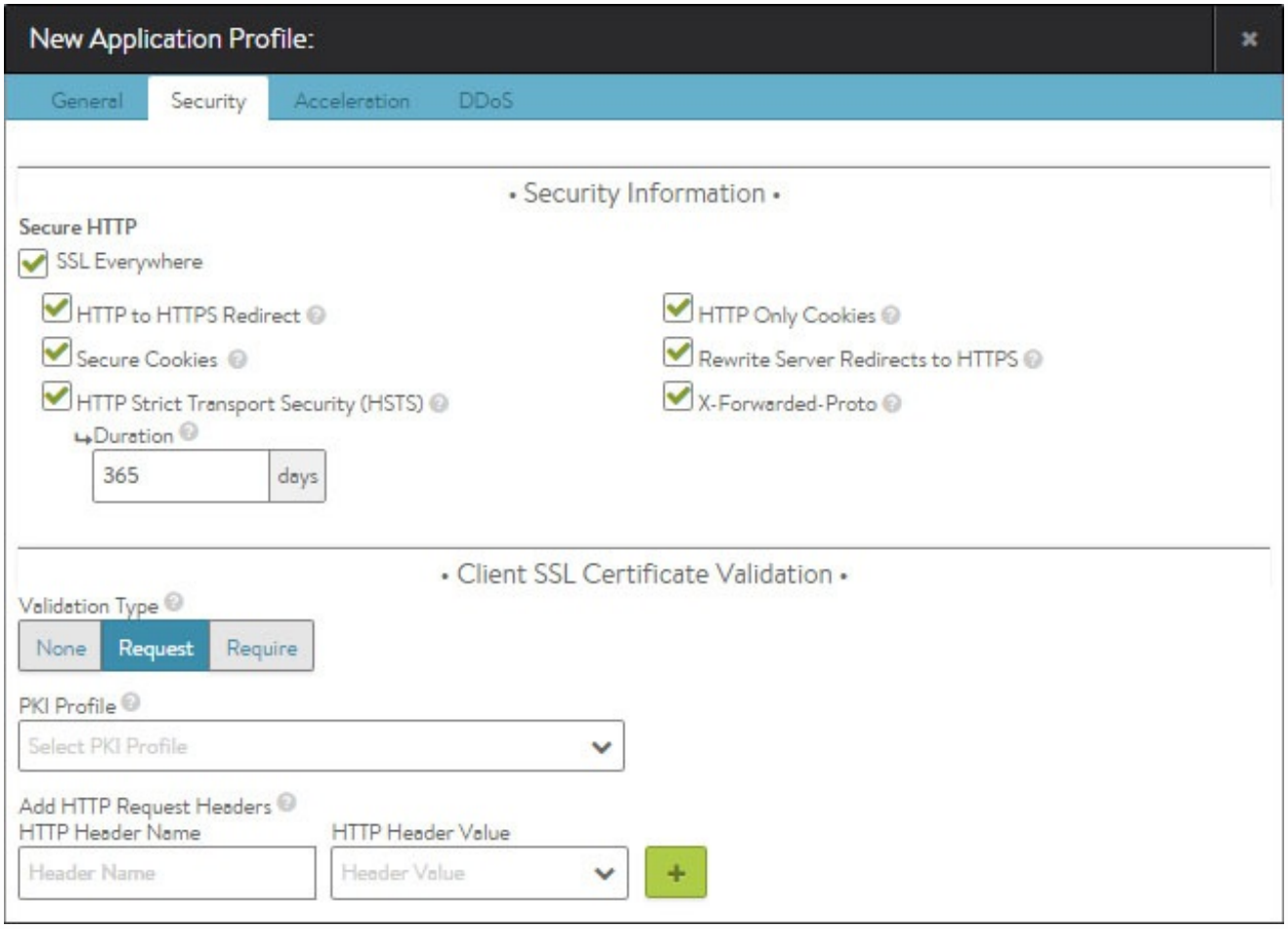
- **Multiplex disabled, Persistence enabled:** Vantage opens a new TCP connection to the server for each connection received from the client. Connections are not shared with other clients. All requests received through all connections from the same client are sent to one server. HTTP client browsers may open many concurrent connections, and the number of client connections will be the same as the number of server connections.

- **Multiplex disabled, Persistence disabled:** Connections between the client and server are

one-to-one. Requests remain on the same connection they began on. Multiple connections from the same client may be distributed among the available servers.

HTTP Security

The Security tab of the HTTP application profile controls the security settings for HTTP applications that are associated with the profile:



Security Information

The HTTP security settings affect how a virtual service should handle HTTPS. If a virtual service is configured only for HTTP, any HTTPS settings in this section will not apply. Only if the virtual service is configured for HTTPS, or HTTP and HTTPS, will the settings take effect.

More granular settings also may be configured usingpolicies or DataScripts.

- **SSL Everywhere:** This option enables all of the following options, which together provide the recommended security for HTTPS traffic.
- **HTTP to HTTPS Redirect:** For a single virtual service configured with both an HTTP service port (SSL disabled) and an HTTPS service port (SSL enabled), this feature will automatically redirect clients from the insecure to the secure port. For instance, clients who type www.avinetworks.com into their browser will automatically be redirected to

https://www.avinetworks.com. If the virtual service does not have both an HTTP and HTTPS service port configured, this feature will not activate. For two virtual services (one with HTTP and another on the same IP address listening to HTTPS), an HTTP request policy must be created to manually redirect the protocol and port.

- **Secure Cookies:** When Vantage is serving as an SSL proxy for the backend servers in the virtual service's pool, Vantage communicates with the client over SSL. However, if Vantage communicates with the backend servers over HTTP (not over SSL), the servers will incorrectly return responses as HTTP. As a result, cookies that should be marked as secure will not be so marked. Enabling secure cookies will mark any server cookies with the Secure flag, which tells clients to send only this cookie to the virtual service over HTTPS. This feature will only activate when applied to a virtual service with SSL/TLS termination enabled.

- **HTTP Strict Transport Security (HSTS):** Strict Transport Security uses a header to inform client browsers that this site should be accessed only over SSL/TLS. This feature is intended to mitigate man-in-the-middle attacks that can force a client's secure SSL/TLS session to connect through insecure HTTP. HSTS has a duration setting that tells clients the SSL/TLS preference should remain in effect for the specified length of time. This setting will activate only on a virtual service that is configured to terminate SSL/TLS.

- **HTTP-only Cookies:** This marks server cookies as HTTPonly, which means the cookies cannot be viewed or used by third parties, including Javascript or other web sites. This feature will activate for any HTTP or terminated HTTPS virtual service.

- **Rewrite Server Redirects to HTTPS:** When a virtual service terminates client SSL/TLS and then passes requests to the server as HTTP, many servers assume that the connection to the client is HTTP. Absolute redirects generated by the server may therefore include the protocol, such as HTTP://www.avinetworks.com. If the server returns a redirect with HTTP in the location header, this feature will rewrite it to HTTPS. Also, if the server returns a redirect for its own IP address, this will be rewritten to the hostname requested by the client. If the server returns redirects for hostnames other than what the client requested, they will not be altered.

  Note: Consider creating an HTTP response policy if greater granularity is required when rewriting redirects. This feature will activate only if the virtual service has both HTTP and HTTPS service ports configured.

- **X-Forwarded-Proto:** Enabling this option makes Vantage insert the X-Forwarded-Proto

header into HTTP requests sent to the server, which informs that server whether the client connected to Vantage over HTTP or HTTPS. This feature activates for any HTTP or HTTPS virtual service.

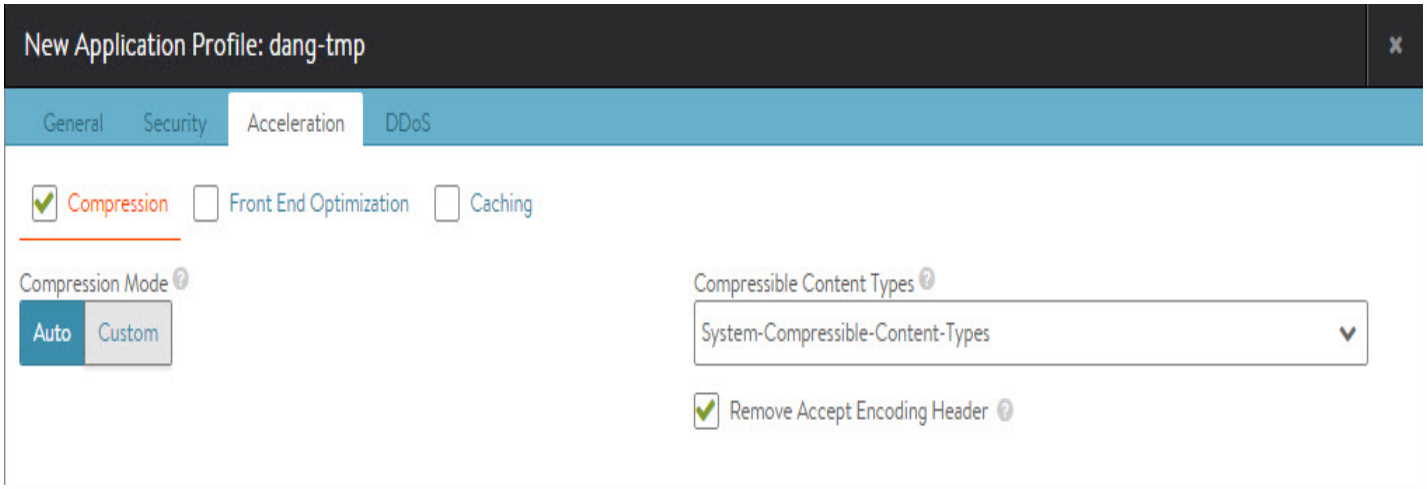
Client SSL Certificate Validation

Vantage can validate the certificates presented by clients, by checking them against a Client Revocation List (CRL). Further options allow passing certificate information to the server through HTTP headers.

- **Validation Type:** Enables client validation based on their SSL certificates.
  - **None:** Disables validation of client certificates.
  - **Request:** This setting expects clients to present a client certificate. If a client does not present a certificate, or if the certificate fails the CRL check, the client connection and requests are still forwarded to the destination server. This allows Vantage to forward the client's certificate to the server in an HTTP header, so that the server may make the final determination to allow or deny the client.
  - **Require:** Vantage requires a certificate to be presented by the client, and the certificate must pass the CRL check. The client certificate, or relevant fields, may still be passed to the server through an HTTP header.
- **PKI Profile:** The Public Key Infrastructure (PKI) profile contains configured certificate authority (CA) and the CRL. A PKI profile is not necessary if validation is set to Request, but is required if validation is set to Require.
- **HTTP Header Name:** Optionally, Vantage may insert the client's certificate, or parts of it, into a new HTTP header to be sent to the server. To insert a header, this field is used to determine the name of the header.
- **HTTP Header Value:** Used with the HTTP Header Name field, the Value field is used to determine the portion of the client certificate to insert into the HTTP header sent to the server. Using the plus icon, additional headers may be inserted. This action may be in addition to any performed by HTTP policies or DataScripts, which could also be used to insert headers in requests sent to the destination servers.

HTTP Acceleration

The Acceleration tab governs the acceleration settings for the HTTP application profile. To edit HTTP acceleration settings:



- Check the Compression checkbox to enable HTTP Gzip compression,, then enter the appropriate compression settings.
- Check the Front End Optimization checkbox to enable SPDY 3.1 protocol support.
- Check the Caching checkbox to enable caching, then enter the appropriate caching settings.

HTTP Compression

The compression option enables HTTP Gzip compression for responses from Vantage to the client. Compression is an HTTP 1.1 standard for reducing the size of text-based data using the Gzip algorithm. The typical compression ratio for HTML, Javascript, CSS, and similar text content types is about 75%, meaning that a 20-KB file may be compressed to 5 KB before being sent across the Internet, thus reducing the transmission time by a similar percentage.

The compression percentage achieved can be viewed using the Client Logs tab of the virtual service. This may require enabling full client logs on the virtual service's Analytics tab to log some or all client requests. The logs will include a field showing the compression percentage with each HTTP response.

Note: It is highly recommended to enable compression in conjunction with caching, which together can dramatically reduce the CPU costs of compressing content. When both compression and caching are enabled, an object such as the index.html file will need to be compressed only one time. After an object is compressed, the compressed object is served out of the cache for subsequent requests. Vantage does not needlessly re-compress the object for every client request. For clients that do not support compression, Vantage also will cache an uncompressed version of the object.

To specify compression settings:

1. Check the Compression checkbox to enable compression. You may only change compression settings after enabling this feature.

2. Select either Auto or Custom, which enables different levels of compression for different clients. For instance, filters can be created to provide aggressive compression levels for slow mobile clients while disabling compression for fast clients from the local intranet. Auto is recommended, to dynamically tune the settings based on clients and available Service Engine CPU resources.
    - **Auto** mode enables Vantage to determine the optimal settings.
    - **Custom** mode allows creation of custom filters that provide more granular control over who should receive what level of compression.
3. **Compressible Content Types** determine which HTTP Content-Types are eligible to be compressed. This field points to a String Group which contains the compressible type list.
4. **Remove Accept Encoding Header** removes the Accept Encoding header, which is sent by HTTP 1.1 clients to indicate they are able to accept compressed content. Removing the header from the request prior to sending the request to the server allows Vantage to ensure the server will not compress the responses. Only Vantage will perform compression.

Custom Compression

To create a custom compression filter:

1. Click **Add New Filter** to create a custom filter.
2. Enter the following:
    - **Filter Name:** Provide a unique name for the filter (optional).
    - **Matching Rules** determine if the client (via Client IP or User Agent string) is eligible to be compressed via the associated Action. If both Client IP and User Agent rules are populated, then both must be true for the compression action to fire.
        - **Client IP Address** allows you to use an IP Group to specify eligible client IP addresses. For example, an IP Group called Intranet that contains a list of all internal IP address ranges. Clearing the Is In button reverses this logic, meaning that any client that is not coming from an internal IP network will match the filter.
        - **User Agent** matches the client's User Agent string against an eligible list contained within a String Group. The User Agent is a header presented by clients indicating the type of browser or device they may be using. The System-Devices-Mobile Group contains a list of HTTP User Agent strings for common mobile browsers.
3. The Action section determines what will happen to clients or requests that meet the Match criteria, specifically the level of HTTP compression that will be used.

- **Aggressive** compression uses Gzip level 6, which will compress text content by about 80% while requiring more CPU resources from both Avi Vantage and the client.
- **Normal** compression uses Gzip level 1, which will compress text content by about 75%, which provides a good mix between compression ratio and the CPU resources consumed by both Avi Vantage and the client.
- **No Compression** disables compression. For clients coming from very fast, high bandwidth and low latency connections, such as within the same data center, compression may actually slow down the transmission time and consume unnecessary CPU resources.

HTTP Front-end Optimization

Client interaction with an HTTP website can be accelerated by enabling the SPDY protocol. Vantage supports SPDY 3.1. While SPDY is officially deprecated and replaced by HTTP/2, it may take many years for all browsers to fully support the newer protocol. SPDY 3.1 is the latest version of the protocol and also is the last version.

SPDY requires clients to connect using HTTP 1.1. If a client presents a header indicating support for SPDY 3.1, Vantage will upgrade the connection to SPDY. HTTP 1.0 clients or 1.1 clients that do not request SPDY will remain connected using normal HTTP 1.0 or 1.1.

Vantage is a full proxy, so clients may negotiate SPDY, even though the servers on the back end are still using HTTP 1.1. No changes to the server environment are required to enable this optimization.

SPDY increases the end-user experience through a number of enhancements. Primarily, it changes the data transfer model from a parallel to a serial model. Rather than allow only a single HTTP request-response exchange at a time per connection, SPDY allows multiple requests and responses simultaneously. This means SPDY opens only a single TCP connection, unlike HTTP 1.1 which may open as many as six TCP connections per client.

HTTP Caching

Avi Vantage can cache HTTP content, thereby enabling faster page load times for clients and reduced workloads for both servers and Vantage. When a server sends a response, such as logo.jpg, Vantage can add the object to its cache and serve it to subsequent clients that request the same object. This can reduce the number of connections and requests sent to the server.

Enabling caching and compression allows Vantage to compress text-based objects and store

both the compressed and original uncompressed versions in the cache. Subsequent requests from clients that support compression will be served from the cache, meaning that Vantage will need not compress every object every time, which greatly reduces the compression workload.

Note: Regardless of the configured caching policy, an object can be cached only if it is eligible for caching. Some objects may not be eligible for caching.

To enable caching, check the Caching checkbox.

The following parameters all are optional:

- **X-Cache:** Vantage will add an HTTP header labeled X-Cache for any response sent to the client that was served from the cache. This header is informational only, and will indicate the object was served from an intermediary cache.
- **Age Header:** Vantage will add a header to the content served from cache that indicates to the client the number of seconds that the object has been in an intermediate cache. For example, if the originating server declared that the object should expire after 10 minutes and it has been in the Vantage cache for 5 minutes, then the client will know that it should only cache the object locally for 5 more minutes.
- **Date Header:** If a date header was not added by the server, then Vantage will add a date header to the object served from its HTTP cache. This header indicates to the client when the object was originally sent by the server to the HTTP cache in Vantage.
- **Cacheable Object Size:** The minimum and maximum size of an object (image, script, and so on) that can be stored in the Vantage HTTP cache, in bytes. Most objects smaller than 100 bytes are web beacons and should not be cached despite being image objects.
- **Cache Expire Time:** An intermediate cache must be able to guarantee that it is not serving stale content. If the server sends headers indicating how long the content can be cached (such as cache control), then Vantage will use those values. If the server does not send expiration timeouts and Vantage is unable to make a strong determination of freshness, then Vantage will store the object for no longer than the duration of time specified by the Cache Expire Time.
- **Heuristic Expire:** If a response object from the server does not include the Cache-Control header but does include an If-Modified-Since header, then Vantage will use this time to calculate the cache-control expiration, which will supersede the Cache Expire Time setting for this object.
- **Cache URL with Query Arguments:** This option allows caching of objects whose URI includes a query argument. Disabling this option prevents caching these objects. When

enabled, the request must match the URI query to be considered a hit. Below are two examples of URIs that include queries. The first example may be a legitimate use case for caching a generic search, while the second may be a unique request posing a security liability to the cache.

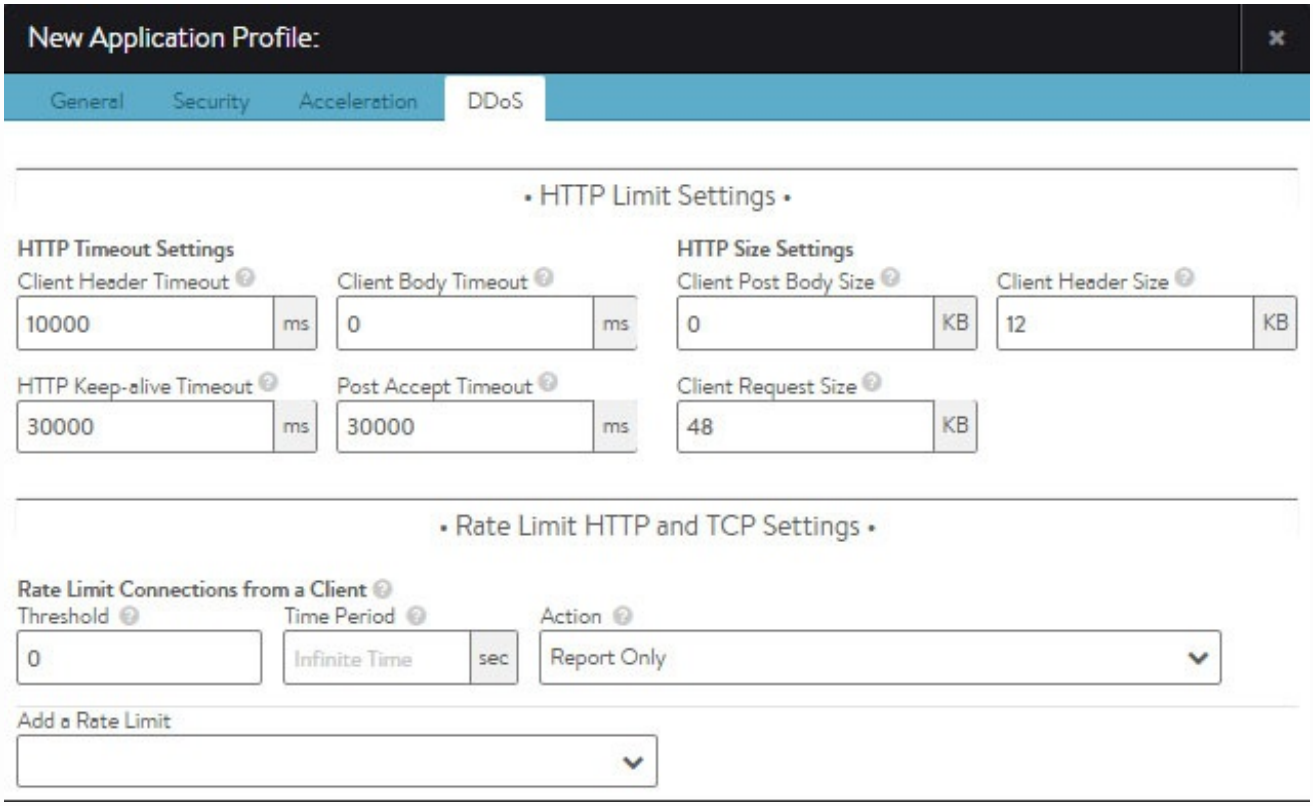
  - www.search.com/search.asp?search=caching
  - www.foo.com/index.html?loginID=User

- **Cacheable Mime Types:** Statically defines a list of cacheable objects. This may be a String Group, such as System-Cacheable-Resource-Types, or a custom comma-separated list of Mime types that Vantage should cache. If no Mime Types are listed in this field, then Vantage will by default assume that any object is eligible for caching.
- **Non-Cacheable Mime Types:** Statically define a list of objects that are not cacheable. This creates a blacklist that is the opposite of the cacheable list.

HTTP DDoS

The Distributed Denial of Service (DDoS) section allows configuration of mitigation controls for HTTP and the underlying TCP protocols. By default, Vantage is configured to protect itself from a number of types of attacks. For instance, if a virtual service is targeted by a SYN flood attack, Vantage will activate SYN cookies to validate clients before opening connections. Many of the options listed below are not quite as straightforward, as bursts of data may be normal for the application. Vantage provides a number of knobs to modify the default behavior to ensure optimal protection.

In addition to the DDoS settings described below, Vantage also can implement connection limits to a virtual service and a pool, configured through the Advanced properties page. Virtual services also may be configured with connection rate limits and burst limits in the Network Security Policies section. Because these settings apply on to an individual virtual service and pool, they are not configured within the profile.



HTTP Limits

The first step in mitigating HTTP-based denial of service attacks is to set parameters for the transfer of headers and requests from clients. Many of these settings protect against variations of HTTP SlowLoris and SlowPOST attacks, in which a client opens a valid connection then very slowly streams the request headers or POSTs a file. This type of attack is intended to overwhelm the server (in this case the Service Engine) by tying up buffers and connections.

Clients that exceed the limits defined below will have that TCP connection reset and a log generated. This does not prevent the client from initiating a new connection and does not interrupt other connections the same client may have open.

- **Client Header Timeout:** Set the maximum length of time the client is allowed for successfully transmitting the complete headers of a request. The default is 10 seconds.
- **HTTP Keep-alive Timeout:** Set the maximum length of time an HTTP 1.0 or 1.1 connection may be idle. This affects only client-to-Vantage interaction. The Vantage-to-server keep-alive is governed through the Connection Multiplex feature.
- **Client Body Timeout:** Set the maximum length of time for the client to send a message body. This usually affects only clients that are POSTing (uploading) objects. The default value of 0 disables this timeout.
- **Post Accept Timeout:** Once a TCP three-way handshake has successfully completed, the client has this much time to send the first byte of the request header. Once the first byte has

been received, this timer is satisfied and the client header timeout (described above) kicks in.

- **Client Post Body Size:** Set the maximum size of the body of a client request. This generally limits the size of a client POST. Setting this value to 0 disables this size limit.
- **Client Request Size:** Set the maximum combined size of all the headers in a client request.
- **Client Header Size:** Set the maximum size of a single header in a client request.

Rate Limits

This section controls the rate at which clients may interact with the site. Each enabled rate limit has three settings:

- **Threshold:** The client has violated the rate limit when the defined threshold of connections, packets, or HTTP requests have occurred within the specified time Period.
- **Time Period:** The client has violated the rate limit when the defined threshold of connections, packets, or HTTP requests have occurred within the specified time Period.
- **Action:** Select the action to perform when a client has exceeded the rate limit. The options will depend on whether the limit is a TCP limit or an HTTP limit.
    - **Report Only:** A log is generated on the virtual server log page. By default, no action is taken. However, this option may be used with an alert to generate an alert action to send a notice to a remote destination or to take action through a ControlScript.
    - **Drop SYN Packets:** For TCP-based limits, silently discard TCP SYNs from the client. Vantage also will generate a log. However, during high volumes of DoS traffic, repetitive logs may be skipped.
    - **Send TCP RST:** Reset client TCP connection attempts. While more graceful than the Drop SYN Packet option, sending a TCP reset does generate extra packets for the reset, versus the Drop SYN Packet option which does not send a client response. Vantage also will generate a log. However, during high volumes of DoS traffic, repetitive logs may be skipped.
    - **Close TCP Connection:** Resets a client TCP connection for an HTTP rate limit violation.
    - **Send HTTP Local Response:** The Service Engine will send an HTTP response directly to the client without forwarding the request on to the server. Select the HTTP status code of the response, and optionally a response page.
    - **Send HTTP Redirect:** Redirect the client to another location.

The following rate limits may be configured.

- **Rate Limit Connections from a Client:** Rate limit all connections made from any single client IP address to the virtual service.
- **Rate Limit Requests from a Client to all URLs:** Rate limit all HTTP requests from any single client IP address to all URLs of the virtual service.
- **Rate Limit Requests from all Clients to a URL:** Rate limit all HTTP requests from all client IP addresses to any single URL.
- **Rate Limit Requests from a Client to a URL:** Rate limit all HTTP requests from any single client IP address to any single URL.
- **Rate Limit Failed Requests from a Client to all URLs:** Rate limit all requests from a client for a specified period of time once the count of failed requests from that client crosses a threshold for that period. Clients are tracked based on their IP address. Requests are deemed failed based on client or server side error status codes, consistent with how Vantage logs and how metrics subsystems mark failed requests.
- **Rate Limit Failed Requests from all Clients to a URL:** Rate limit all requests to a URI for a specified period of time once the count of failed requests to that URI crosses a threshold for that period. Requests are deemed failed based on client- or server-side error status codes, consistent with how Vantage logs and metrics subsystems mark failed requests.
- **Rate Limit Failed Requests from a Client to a URL:** Rate limit all requests from a client to a URI for a specified period of time once the count of failed requests from that client to the URI crosses a threshold for that period. Requests are deemed failed based on client- or server-side error status codes, consistent with how Vantage logs and metrics subsystems mark failed requests.
- **Rate Limit Scans from a Client to all URLs:** Automatically track clients and classify them into three groups: Good, Bad, and Unknown. Clients are tracked based on their IP address. Clients are added to the Good group when the Vantage scan detection system builds history of requests from the clients that complete successfully. Clients are added to the Unknown group when there is insufficient history about them. Clients with a history of failed requests are added to the Bad group and their requests are rate limited with stricter thresholds than the Unknown clients group. The Vantage scan detection system automatically tunes itself so that the Good, Bad, and Unknown client-IP group members change dynamically with changes in traffic patterns through Vantage. In other words, if a change to the website causes mass failures (such as 404 errors) for most customers, Vantage adapts and does not mark all clients as attempting to scan the site.
- **Rate Limit Scans from all Clients to all URLs:** Similar to the previous limit, but restricts the

scanning from all clients as a single entity rather than individually. Once a limit is collectively reached by all clients, any client that sends the next failed request will be reset.

DNS Profile

The DNS application profile allows Vantage to decode the DNS protocol. This profile will set the virtual service to port 53, and the network profile to UDP with per-packet parsing.

L4 Profile

The L4 Profile is used for any virtual service that does not require application-layer proxying.

Note: Using an L4 profile is equivalent to setting the virtual service's application profile to 'none'.

Rate limits may be placed on the number of TCP connections or UDP packets that may be made to the virtual service from a single client IP address.

- **Threshold:** The client has violated the rate limit when the defined threshold of connections (TCP) or packets (UDP) is reached within the specified time period.
- **Time Period:** The client has violated the rate limit when the defined threshold of connections (TCP) or packets (UDP) is reached within the specified time period.
- **Action:** Select the action to perform when a client has exceeded the rate limit.
    - **Report Only:** A log is generated in the virtual service logs page. By default, no action is taken. However, this option may be used with an alert to generate an alert action to send a notice to a remote destination or to take action using a ControlScript.
    - **Drop SYN Packets:** For TCP-based limits, silently discard TCP SYNs from the client. Vantage also will generate a log. However, during high volumes of DoS traffic, repetitive logs may be skipped.
    - **Send TCP RST:** Reset client TCP connection attempts. While more graceful than the Drop SYN Packet option, sending a TCP reset does generate extra packets for the reset, versus the Drop SYN Packet option which does not send a client response. Vantage also will generate a log. However, during high volumes of DoS traffic, repetitive logs may be skipped.

Syslog Profile

The Syslog application profile allows Vantage to decode the Syslog protocol. This profile will set the virtual service to understanding Syslog, and the network profile to UDP with per-stream parsing.

TCP/UDP Profile

A TCP/UDP Profile determines the type and settings of the network protocol that a subscribing virtual service will use. It sets a number of parameters, such as whether the virtual service is a TCP proxy versus a pass-through via fast path. A virtual service can have both TCP and UDP enabled, which is useful for protocols such as DNS or Syslog.

Avi Vantage will rewrite the client IP address before sending any TCP connection to the server, regardless which type of TCP profile is used by a virtual service. Similarly, the destination address will be rewritten from the virtual service IP address to the IP address of the server. The server will always see the source IP address of the Service Engine. UDP profiles have an option to disable Service Engine Source NAT.

For the UDP and TCP fast path modes, connections occur directly between the client and server, even though the IP Address field of the packet has been altered.

For HTTP applications, Avi Vantage can insert the client's original IP address via XFF into an HTTP header sent to the server.

This section contains the following sub-chapters.

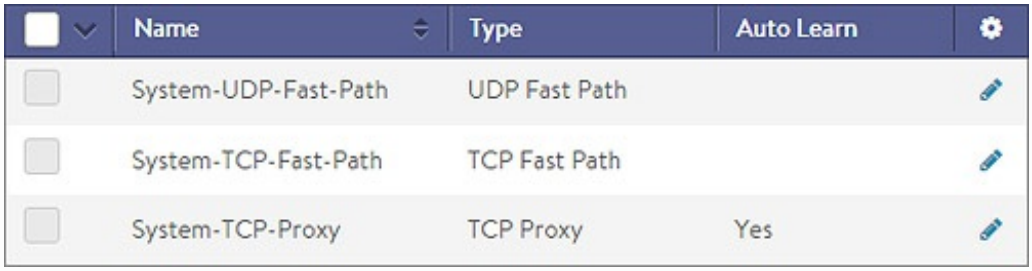
- TCP Proxy
- TCP Fast Path
- UDP Fast Path

TCP/UDP Profiles Settings

Select Templates > Profiles > TCP/UDP to open the TCP/UDP Profiles tab. This tab includes the following functions:

- **Search:** Search across the list of objects.
- **Create:** Opens the New TCP/UDP Profile popup.
- **Edit:** Opens the Edit TCP/UDP Profile popup.
- **Delete:** A TCP/UDP profile may only be deleted if it is not currently assigned to a Virtual Service. An error message will indicate the Virtual Service referencing the profile. The default system Profiles cannot be deleted.

The table on this tab provides the following information for each TCP/UDP profile:



- **Name:** Name of the profile.
- **Type:** Type of TCP/UDP profile, which can be one of the following:
  - **TCP Proxy:** This profile terminates client connections to the Virtual Service and then open a new TCP connection to the destination server. Each connection will negotiate the optimal TCP settings for the connecting device. For example, a client may connect with a 1400-byte MTU while the server can still send data to Avi Vantage in 1500-byte MTUs. In this case, Avi Vantage will buffer the server's responses and send them back to the client separately. If the client connection drops a packet, then Avi Vantage will handle retransmission, as the server may have already finished the transmission and moved on to handling the next client request. See TCP Proxy.
  - **TCP Fast Path:** Upon receiving a TCP SYN from the client, Avi Vantage will make a load balancing decision and forward the SYN and all subsequent packets directly the server. The client's source IP address will still be translated to the Service Engine's IP address for the server network to ensure return path routing. The client to server communication occurs over a single TCP connection, using the parameters negotiated between client and server. See TCP Fast Path.
  - **UDP Fast Path:** UDP is connectionless, meaning that packets are directly forwarded to the load balanced server. The load balancing decision is made on the first packet from the client, and the source IP address is still changed to the Service Engine's IP address. See UDP Fast Path.
- **Auto Learn:** While in the default Auto Learn mode, the TCP/UDP profile will dynamically adjust settings based on the application type assigned to the Virtual Service. Disabling Auto Learn uses the parameters statically defined within the profile.

Create TCP/UDP Profile

To create or edit a TCP/UDP profile:

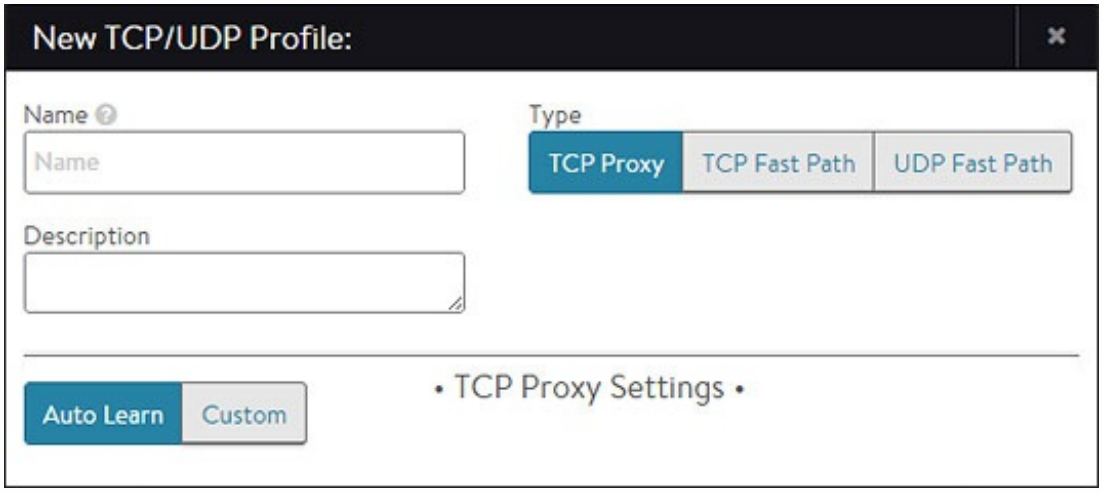
1. **Name:** Enter a unique name for the profile.
2. **Type:** Select the type of TCP/UDP Profile:

- **TCP Proxy:** This option terminates client TCP connections at the Virtual Service. A new TCP connection will then be used for the connection between the Service Engine and the server. This type is required in order to apply an HTTP application profile to the virtual service. See TCP Proxy.
- **TCP Fast Path:** This option will allow a client to connect directly with the server through Avi Vantage. This minimizes the workload for Avi Vantage, but also reduces security, visibility, acceleration, and application availability. See TCP Fast Path.
- **UDP Fast Path:** This option enables UDP for the Virtual Service. SeeUDP Fast Path.

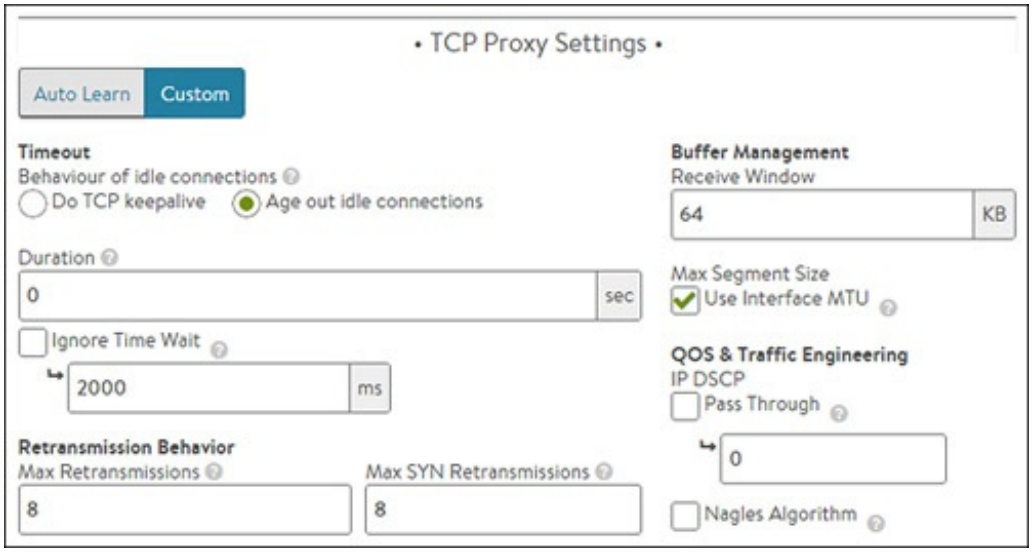
3. Click Save to save your changes and return to the TCP/UDP Profiles tab.

TCP Proxy

Enabling TCP Proxy causes Avi Vantage to terminate an inbound connection from a client. Any application data from the client that is destined for a server is forwarded to that server over a new TCP connection. Separating (or proxying) the client to server connections allows Avi Vantage to provide enhanced security, such as TCP protocol sanitization or DoS mitigation. It also provides better client and server performance, such as maximizing client and server TCP MSS or Window sizes independently and buffering server responses. You must use a TCP/UDP profile with the type set to Proxy for application profiles such as HTTP.



Select TCP Proxy in the Create/Edit TCP/UDP Profile popup and select either Auto Learn or Custom. With Auto Learn, Avi Vantage will dynamically adjust the following parameters based on the Virtual Server application it is assigned to. This option is selected by default and is the easiest way to ensure optimal performance. Select Custom to manually configure the following parameters:



- **Timeout:** Controls the behavior of idle connections, as follows:
  - **Do TCP keepalive:** Send a periodic keep-alive signal to the client that will keep the current connection open.
  - **Age out idle connections:** Terminate idle connections that have no keep-alive signal from the client, as specified by the Duration field.
  - **Idle Duration:** Idle time, in seconds, before the TCP connection is eligible to be proactively closed by Avi Vantage. The timer resets when any packet is sent or received. Setting this value higher may be appropriate for long-lived connections that do not use keepalive packets; however, higher settings may also increase the vulnerability of Avi Vantage to Denial of Service attacks, as the system will not proactively close out idle connections. You may either specify a range of 60-1800 seconds, or 0 for an infinite timeout.
  - **Ignore Time Wait:** When a connection between either a Service Engine and a client or the Service Engine and a server is closed, the unique client or server IP:Port + Service Engine IP:Port (called a 4-tuple) is placed in a TIME_WAIT state for a period of time. This 4-tuple cannot be reused until it is clear that there are no more delayed packets on the network that are still in flight or that have yet to be delivered. The Time Wait value defines the timeout period before this 4-tuple may be reused. This may either be a value between 500 – 2000 (default) ms or enable the Ignore Time Wait checkbox to allow Avi Vantage to immediately reopen the 4-tuple connection if it receives a SYN packet from the remote IP that matches the same 4-tuple.
- **Retransmission Behavior:**
  - **Max Retransmissions:** Maximum number of attempts (3-8) to retransmit packets before giving up and closing the connection.

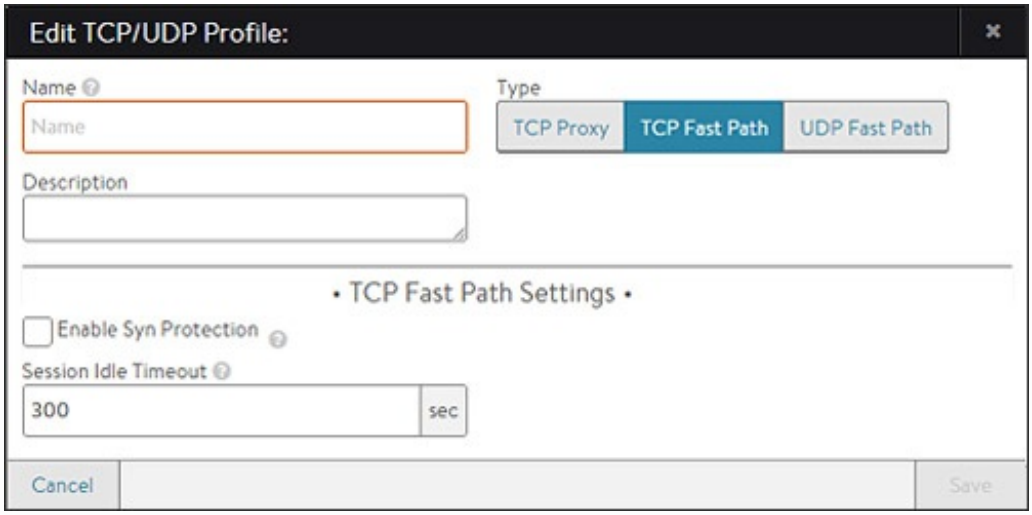
- **Max SYN Retransmissions:** Maximum number of attempts (3-8) to retransmit a SYN packet before giving up.

- **Buffer Management:**
  - **Receive Window:** The Receive Window informs the sender how much data Avi Vantage can buffer (32-256KB) before sending a TCP acknowledgement.
  - **Max Segment Size (MSS):** This may be calculated by using the maximum transmission unit (MTU) length for a network interface. The MSS determines the largest size of data that may be safely inserted into a TCP packet. In some environments, the MSS should be smaller than the MTU. For example, traffic between Avi Vantage and a client that is traversing a site-to-site VPN may require some space reserved for padding with encryption data. Select either:
    - **Use Interface MTU:** Sets the MSS to the MTU size of the network interface.
    - **Custom value:** May be a range between 512–9000 kilobytes.

- **QOS & Traffic Engineering:**
  - **IP DSCP:** Allows Avi Vantage to either pass an existing Differentiated Services Code Point (DSCP) parameter or specify a custom number. DSCP is an 8-bit field in the TCP header that can be used for classifying traffic in a manner similar to the deprecated TCP TOS field.
  - **Nagles Algorithm:** Attempts to reduce latency by combining small packets into a smaller quantity of larger packets before sending. This reduces the impact of network latency by reducing the number of round trip times required due to TCP acknowledgements. This option could adversely impact real time protocols, particularly SSH and Telnet. For example, typing in a telnet session many not reflect any text back until either a user types 1500 characters (enough to fill a typical packet) or enough time has elapsed and the packet is sent half-full.

TCP Fast Path

A TCP Fast Path profile does not proxy TCP connections; rather, it directly connects clients to the destination server and translates the client’s destination Virtual Service address with the chosen destination server’s IP address. The client’s source IP address is still translated to the Service Engine address to ensure that server response traffic returns symmetrically.



In this mode, network negotiation such as MSS or TCP windows occurs between the client and the server. TCP fast path is less work (CPU overhead) for Vantage than TCP proxy mode; however, it also results in greater network overhead and latency for both the client and server. Further, a virtual service using TCP fast path may not use application layer profiles such as HTTP.

This profile type has the following settings:

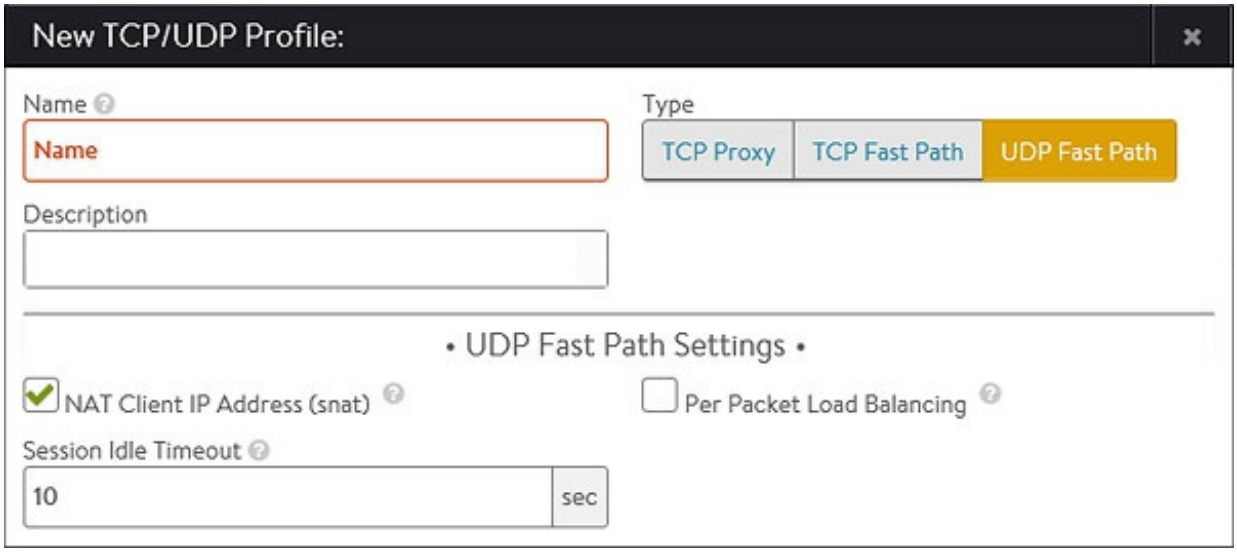
- **Enable SYN Protection:** When disabled, Avi Vantage performs load balancing based on the initial client SYN packet. The SYN is forwarded on to the server and Avi Vantage merely forwards the packets between client and server, which leaves servers vulnerable to SYN flood attacks from spoofed IP addresses. When enabled, Avi Vantage will proxy the initial TCP three-way handshake with the client to validate the client is not a spoofed source IP address. Once the three-way handshake has been established, Avi Vantage will replay the handshake on the server side. After the client and server are connected, Avi Vantage will drop back into pass through (fast path) mode. This process is sometimes referred to as delayed binding.

snoC :etoN

- **Session Idle Timeout:** Idle flows will terminate (time out) after the specified time period. Avi Vantage will issue a TCP reset to both the client and the server.

UDP Fast Path

The UDP Fast Path profile allows a Virtual Service to support UDP. Avi Vantage will translate the client's destination Virtual Service address to the destination server and rewrite the client's source IP address to the Service Engine's address when forwarding the packet to the server. This ensures that server response traffic traverses symmetrically through the original SE.



This profile type uses the following settings:

- **NAT Client IP Address:** By default, Avi Vantage will translate the client's source IP address to an IP address of the Avi Service Engine. This may be disabled for connectionless protocols which do not require server response traffic to traverse back through the same Service Engine. For example, a syslog server will silently accept packets without responding. Therefore, there is no need to ensure response packets route through the same SE. When SNAT is disabled, it is recommended to ensure the Session Idle Timeout is kept to a lower value.
- **Per Packet Load Balancing:** By default, Avi Vantage treats a stream of UDP packets from the same client IP:Port as a session, making a single load balancing decision and sending subsequent packets to the same destination server. For some application protocols, each packet should be treated as a separate session that can be uniquely load balanced to a different server. DNS is one example where enabling Per Packet Load Balancing causes Avi Vantage to treat each packet as an individual session or request.
- **Session Idle Timeout:** Idle UDP flows will terminate (time out) after the specified time period. Subsequent UDP packets could be load balanced to a new server unless a persistence profile is applied.

Persistence Profile

A persistence profile governs the settings that will force a client to stay connected to the same server for a specified duration of time. By default, load balancing may send a client to a different server every time the client connects with a virtual service. This guarantees the client will reconnect to the same server every time when they connect to a virtual service, disconnect, and then reconnect to the same virtual service. Enabling a persistence profile ensures the client will

reconnect to the same server every time, or at least for a desired duration of time. Persistent connections are critical for most servers that maintain client session information locally.

Persistence is an optional profile that is attached to a pool.

Persistence Profile Settings

Select Templates > Profiles > Persistence to open the Persistence Profiles tab.

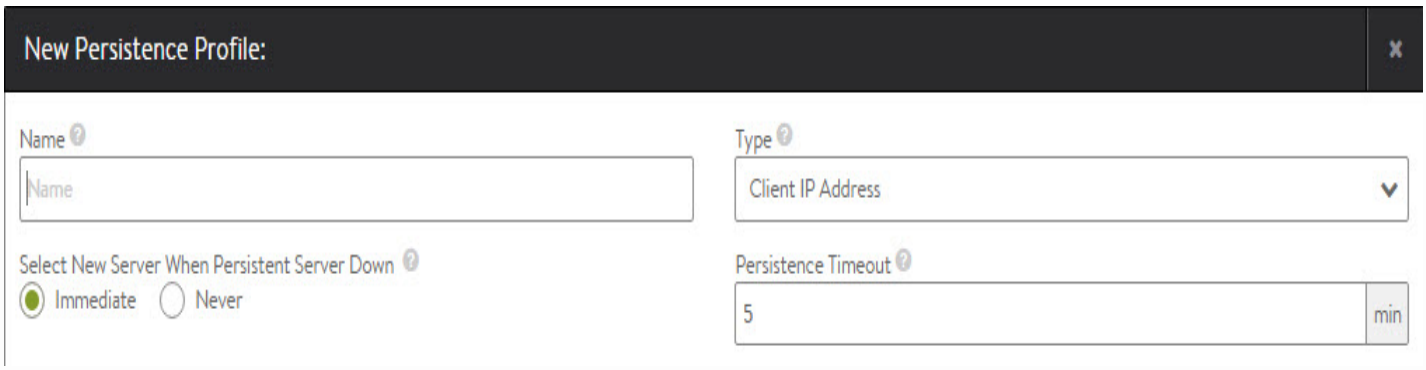
- **Search:** Search across the list of objects.
- **Create:** Opens the New Persistence Profile popup.
- **Edit:** Opens the Edit Persistence Profile popup.
- **Delete:** A profile may only be deleted if it is not currently assigned to a virtual service. An error message will indicate the virtual service referencing the profile. The default system profiles can be edited, but not deleted.

The table on this tab provides the following information for each persistence profile:

- **Persistence Name:** Name of the profile.
- **Type:** Full descriptions of each type of persistence are described in the next section. The types of persistence can be one of the following:
  - Client IP Address
  - HTTP Cookie
  - App Cookie
  - HTTP Custom Header
  - TLS

Create Persistence Profile

The New Persistence Profile and Edit Persistence Profile popups share the same interface.



To create or edit a Persistence Profile:

- **Name:** Enter a unique name for the Persistence Profile in the Name field.
- **Type:** Select the persistence type using the Type pull-down menu. The available options are:
  - **Client IP Address:** Avi Vantage will record the client's source IP address in a table for the duration of the Persistence Timeout for this profile. While the IP remains in the table, any new connection by the user will be sent to the same server. The Client IP Address persistence table is stored in memory on the Service Engine, and is automatically mirrored to the Controller and all other Service Engines that support this virtual service.
    - **Persistence Timeout:** Enter the number of minutes to preserve the client's IP address in the Persistence Timeout field. Entering 0 disables persistence and allows new connections to be load balanced to a new server immediately. The timeout begins counting down when all connections from the same source IP address to the virtual service are closed. This field is applicable to Client IP persistence only.
  - **HTTP Cookie:** Applicable to virtual services with an attached HTTP application profile. Avi Vantage inserts a cookie into outgoing HTTP responses and reads the cookie on incoming requests. The cookie is session based, meaning that the cookie persistence remains valid as long as the client keeps their browser open. Closing the browser removes the cookie stored by the client browser, thereby tearing down both the connections and the persistence. Cookies uniquely identify each client, which is useful if multiple users are accessing the virtual service from the same IP address. Clients store the persistence information, so it does not consume memory on the Service Engine. The cookie is transparent to clients and will be named AVI_COOKIE by default.
    - **HTTP Cookie Name:** By default, the cookie name is AVI_COOKIE unless this field is populated with an alternate name.
  - **Custom HTTP Header:** This method allows an HTTP header to be specified for persistence. The Service Engine will inspect the value of the defined header, and will match the value against a statically assigned header field for each server in the pool. If

there is a match, the client will be persisted. The server's header field is configured in the Pool's edit server page, where new servers are added.

        - **Header Name:** Specify the HTTP header whose value will be used for the persistence lookup.
    - **App Cookie:** Rather than have Avi insert a new cookie for persistence, Avi will use an existing cookie that has been inserted by the server. If the cookie does not exist, Avi will look for a URI query of the same name and will persist on that value. Typically this persistence will be performed on a ASP or Java session ID.
    - **TLS:** Applicable to virtual services that are terminating SSL or TLS. This method embeds user persistence information within the ticket field of a TLS session. Clients negotiating with older SSL v3 will use a variation that inserts the persistence information into the SSL Session ID. Avi Vantage does not allow clients to renegotiate the session, which is more secure and also ensures that Avi Vantage can maintain the persistence as it controls if and when the Session ID is renegotiated and recreated.
- **Select New Server When Persistent Server Down:** Determine how this profile will handle a condition when the Health Monitor marks a server as down while Avi Vantage is still persisting clients to it.
    - **Immediate:** Avi Vantage will immediately select a new server to replace the one that has gone down and switch the persistence entry to the new server.
    - **Never:** No replacement server will be selected. Persistent entries will be required to expire normally based upon the persistence type.

Health Monitor Profile

Avi Vantage must validate whether servers are working correctly and are able to accommodate additional workloads before load balancing a client to a particular server. Health monitors perform this function by either actively sending a synthetic transaction to a server or by passively monitoring client experience with the server. Vantage sends active health monitors on a periodic basis that originate from the Service Engines assigned to the application's virtual service.

The health monitors are attached to the pool for the virtual service. When scaling out a virtual service across multiple Service Engines, servers will receive active health queries from each Service Engine.

A pool that is not attached to a virtual service is considered to be in an unused state and will therefore not send health monitors to its servers. A pool may have both multiple actively

concurrent health monitors (such as Ping, TCP, and HTTP), as well as a passive monitor. All active health monitors must be successful for the server to be marked up.

Only active health monitors may be edited. The passive monitor has no settings.

- Health Monitors Tab
- Create/Edit a Health Monitor
- Passive Health Monitor

Health Monitor Settings

Select Templates > Profiles > Health Monitor to open the Health Monitor tab. This tab includes the following functions:

- **Search:** Search across the list of objects.
- **Create:** Opens the New Health Monitor Profile popup.
- **Edit:** Opens the Edit Health Monitor Profile popup.
- **Delete:** A Profile may only be deleted if it is not currently assigned to a Virtual Service. An error message will indicate the Virtual Service referencing the Profile. The default system Profiles can be edited, but not deleted.

The table on this tab provides the following information for each Health Monitor Profile:

Name	Type	Send Interval	Receive Tim...	Successful...	Failed Checks	
System-Ping	Ping	10 sec	4 sec	2	2	
System-TCP	TCP	10 sec	4 sec	2	2	
System-UDP	UDP	4 sec	2 sec	2	2	

- **Health Monitor:** Name of the Health Monitor.
- **Type:** Type of Health Monitor, which can be one of the following:
  - **DNS:** Validate the health of responses from DNS servers.
  - **External:** Use a custom script to validate the health of a diverse array of applications.
  - **HTTP:** Validate the health of HTTP web servers.

- **HTTPS:** Validate the health of HTTPS web servers when the connection between Avi Vantage and the server is SSL/TLS encrypted.
- **Ping:** An ICMP ping can be used to monitor any server that responds to pings. This is a lightweight monitor, but it does not validate application health.
- **TCP:** Validate TCP applications via simple TCP request/response data.
- **UDP:** Validate UDP applications via simple UDP request/response data.

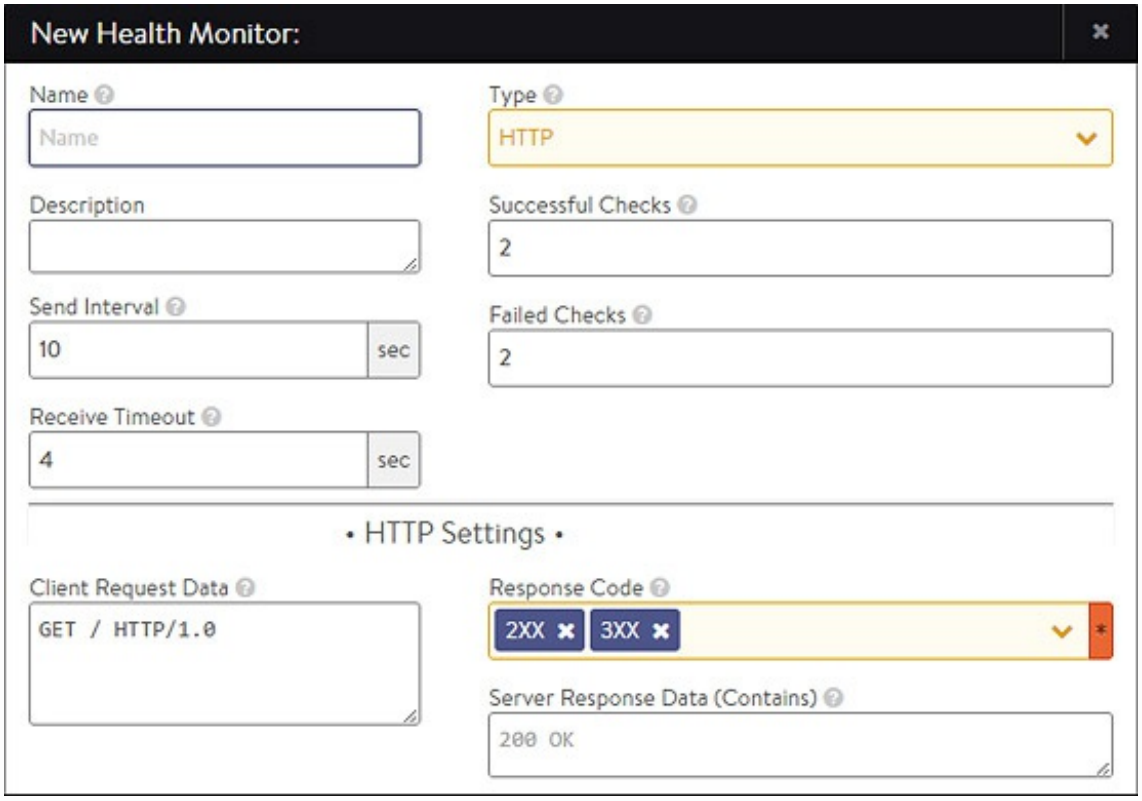
- **Send Interval:** Frequency at which the Health Monitor initiates a server check, in seconds.
- **Receive Timeout:** Maximum amount of time before the server must return a valid response to the Health Monitor, in seconds.
- **Successful Checks:** Number of consecutive health checks that must succeed before Avi Vantage marks a down server as being back up.
- **Failed Checks:** Number of consecutive health checks that must fail before Avi Vantage marks an up server as being down.

Active Health Monitor

Active Health Monitors send proactive queries to servers, synthetically mimicking a client. Send and receive timeout intervals may be defined, which statically determines the server response as successful or failed. Active Health Monitors originate from the Service Engines assigned to the application's Virtual Service. Each SE must be able to send monitors to the servers, which ensures there are no routing or intermediate networking issues that might prevent access to a server from all active Service Engines.

The Health Monitors are attached to the Pool for the Virtual Service. A Pool that is not attached to a Virtual Service will not send health monitors.

The New Health Monitor and Edit Health Monitor popups share the same interface.



To create or edit a Health Monitor enter the following information:

- **Name:** Enter a unique name for the Health Monitor.
- **Send Interval:** Determines how frequently the Health Monitor initiates an active check of a server, in seconds. The minimum frequency is 2 seconds, and the maximum is 3600.
- **Receive Timeout:** Maximum amount of time before the server must return a valid response to the Health Monitor, in seconds. The minimum value is 1 second, and the maximum is the shorter of either 300 seconds or the Send Interval value minus 1 second. If the status of a server continually flips between up and down, this may indicate that the Receive Timeout is too aggressive for the server.
- **Successful Checks:** Number of consecutive health checks that must succeed before Avi Vantage marks a down server as up. The minimum is 1, and the maximum is 50.
- **Failed Checks:** Number of consecutive health checks that must fail before Avi Vantage marks an up server as down. The minimum is 1, and the maximum is 50.
- **Type:** Select the type of Health Monitor and entering the appropriate options:
    - DNS Monitor
    - External Monitor
    - HTTP Monitor

- HTTPS Monitor
- Ping Monitor
- TCP Monitor
- UDP Monitor

DNS Monitor

A DNS Monitor queries Name Servers for an A record and matches the resolved response against an expected IP address.

- **Request Name:** Resource record to be queried, such as www.avinetworks.com.
- **Response Matches:**
  - **Any Thing:** Any DNS answer from the server will be successful, even an empty answer.
  - **Any Type:** The DNS response must contain at least one non-empty answer.
  - **Query Type:** The response must have at least one answer of which the resource record type matches the query type.
- **Response Code:**
  - **Anything:** Ignore the Response Code and any other errors.
  - **No Error:** DNS response has no error as the Response Code.
- **Response String:** IP address in the dotted-decimal notation to match in the resource record sections of the DNS response. The DNS response must contain this IP address to be considered successful.

External Monitor

This Monitor type allows scripts to be written to provide highly customized and granular health checks. The scripts may be Linux shell, Python, or Perl, which can be used to execute wget, netcat, curl, snmpget, or dig. External Monitors have constrained access to resources, such as CPU and memory, so as to not affect normal functioning of Avi Vantage. As with any custom scripting, thoroughly validate the long term stability of the implemented script. Errors generated from the script will be included in the output of the Administrator > Events log.

- **Script Code:** Upload the script via copy/paste or uploading the file.
- **Script Parameters:** Enter any optional arguments to apply. These strings are passed in as

arguments to the script, such as $1 = server IP, $2 = server port.

- **Script Variables:** Custom environment variables may be fed into the script to allow for simplified reusability. For example, a script that authenticates to the server may have a variable set to USER=test.

HTTP Monitor

This Monitor type sends a request to a web server and validates either the HTTP response code or the HTML response data.

- **Client Request Data:** Send an HTTP request to the web server. Avi Vantage does not validate the request syntax, as different servers may support unique request syntax. A basic HTTP request might look like GET /index.html HTTP/1.0.
  - Any method may be used, though GET, POST and HEAD are the most common for monitoring. If no method is specified Avi Vantage will use GET.
  - The Path may include the URI and Query, such as `/index.htm?user=test` . If no Path is specified, Avi Vantage will use `/` .
  - The HTTP version can be 1.0 or 1.1. If no version is specified, Avi Vantage will use 1.0. When using 1.1, web servers expect a Host header to be included with the request, such as GET / index.htm HTTP/1.1 Host: www.avinetworks.com.
- **Response Code:** A successful HTTP monitor requires either the Response Code, the Server Response Data, or both fields to be populated. The Response Code expects the server to return a response code within the specified range. For a GET request, a server should usually return a 200, 301 or 302. For a HEAD request, the server will typically return a 304. A response code by itself does not validate the server's response content, just the status.
- **Server Response Data:** Enter a snippet of content from the server's HTTP response by copying and pasting text from either the source HTML or the web page of the server. Avi Vantage inspects raw HTML data and not rendered web pages. For example, Avi Vantage does not follow HTTP redirects and will compare the redirect response with the defined Server Response string, while a browser will show the redirected page. The Server Response content is matched against the first 2KB of data returned from the server, including both headers and content/body. The Server Response Data can also be used to search for a specific response code, such as 200 OK.

  When both Response Code and Server Response Data are populated, both must be true for the health check to pass.
- **Health Monitor Port:** Specify a port that should be used for the health check. When this

setting is blank, the default port configured for the server will be used. When it is specified, clients may be directed to a different port than what is monitored. For instance, a server at HTTP port 80 may have two health monitors attached, one for HTTP default port, and a second for HTTPS specifically on port 443. If both health monitors pass, the server can receive traffic on HTTP port 80. This ensures clients can input items in their shopping cart and later purchase those items over SSL on 443.

HTTPS Monitor

This Monitor type can be used to validate the health of HTTPS encrypted web servers. Use this monitor when Avi Vantage is either passing SSL encrypted traffic directly from clients to servers, or Avi Vantage is providing SSL encryption between itself and the servers. This monitor is the same as the HTTP Monitor.

- **Client Request Data:** Send an HTTP request to the web server. Avi Vantage does not validate the request, as different servers may support unique request syntax. A basic HTTP request might look like GET /index.html HTTP/1.0.
    - Any method may be used, though GET, POST and HEAD are the most common for monitoring. If no method is defined, Vantage will use GET.
    - The path may include the URI and query, such as `/index.htm?user=test` . If no path is specified, Vantage will use `/` .
    - The HTTP version can be 1.0 or 1.1. If no version is specified, Vantage will use 1.0. When using 1.1, web servers may expect a Host header to be included with the request, such as GET / index.htm HTTP/1.1 Host: www.site.com.
- **Response Code:** A successful HTTP monitor requires either the Response Code, the Server Response Data, or both fields to be populated. The Response Code expects the server to return a response code within the specified range. For a GET request, a server should usually return a 200, 301 or 302. For a HEAD request, the server will typically return a 304. A response code by itself does not validate the server's response content, just the status.
- **Server Response Data:** Enter a snippet of content from the server's HTTP response by copying and pasting text from either the source HTML or the web page of the server. Avi Vantage inspects raw HTML data and not rendered web pages. For example, Avi Vantage does not follow HTTP redirects and will compare the redirect response with the defined Server Response string, while a browser will show the redirected page. The Server Response content is matched against the first 2KB of data returned from the server, including both headers and content/body. The Server Response Data can also be used to search for a specific response code, such as 200 OK.When both Response Code and Server Response

Data are populated, both must be true for the health check to pass.

- **Health Monitor Port:** Specify a port that should be used for the health check. When this setting is blank, the default port configured for the server will be used. When it is specified, clients may be directed to a different port than what is monitored. For instance, a server at HTTP port 80 may have two health monitors attached, one for HTTP default port, and a second for HTTPS specifically on port 443. If both health monitors pass, the server can receive traffic on HTTP port 80. This ensures clients can input items in their shopping cart and later purchase those items over SSL on 443.

Ping Monitor

Avi Vantage will send an ICMP ping to the server. This Monitor type is generally very fast and lightweight for both Avi Vantage and the server; however, it's not uncommon for ping to drop a packet and fail. It is important to ensure the Failed Checks field is set to at least 2. This Monitor type does not test the health of the application, so it generally works best when applied in conjunction with an application specific monitor for the Pool.

TCP Monitor

For any TCP application, this Monitor will wait for the TCP connection establishment, send the request string, and then wait for the server to respond with the expected content. If no client request and server response are configured, the health check will pass once a TCP connection is successfully established.

- **Client Request Data:** Enter an appropriate send string applicable for the server application. This request will be sent to the server immediately after completing the TCP three-way handshake.
- **Server Response Data:** Enter the expected response from the server. Avi Vantage checks to see if the Server Response data is contained within the first 2KB of data returned from the server.
- **Health Monitor Port:** Specify a port that should be used for the health check. When this setting is blank, the default port configured for the server will be used. When it is specified, clients may be directed to a different port than what is monitored.

UDP Monitor

Send the request data to the server, then match the server's response against the expected response data.

- **Client Request Data:** Enter an appropriate send string applicable for the server application.

- **Server Response Data:** Enter the expected response from the server. Avi Vantage checks to see if the Server Response data is contained within the first 2KB of data returned from the server.
- **Health Monitor Port:** Specify a port that should be used for the health check. When this setting is blank, the default port configured for the server will be used. When it is specified, clients may be directed to a different port than what is monitored.

Passive Health Monitor

While active Health Monitors provide a binary good/bad analysis of server health, passive Health Monitors provide a more subtle check by attempting to understand and react to the client-to-server interaction. An active Health Monitor sends a periodic check to the servers that mimics an end user transaction with a synthetic request, then statically validates the response against an expected string. Passive Health Monitors do not send a check to the servers. Instead, Avi Vantage monitors end user interaction with the servers. If a server is quickly responding with valid responses (such as HTTP 200), then all is well; however, if the server is sending back errors (such as TCP resets or HTTP 5xx errors), the server is assumed to have errors. Errors are defined by the Analytics Profile attached to the Virtual Service. The Analytics profile also defines the threshold for response time before a server is considered responding slowly.

With active Health Monitors, Avi Vantage will mark a server down after the specified number of consecutive failures and will no longer send new connections or requests until that server is able to correctly pass the periodic active health Monitors.

With passive Health Monitors, server failures will not cause Avi Vantage to mark that server as down. Rather, the passive Health Monitor will reduce the number of connections or requests sent to the server relative to the other servers in the Pool by about 75%. Further failures may increase this percentage. After a server is satisfactorily handling the reduced load, it will once again be sent normal volumes of traffic.

Analytics Profile

Avi Vantage relies extensively on Analytics throughout the system to determine the health of applications based on expectations of what a typical user experience should be. Since each application is different, it may be necessary to modify the Analytics Profile to set the threshold for satisfactory client experience or omit certain errors from being counted against the application health, such as prompting a user to log in to a site via an HTTP 401 response code.

Analytics Settings

Select Templates > Profiles > Analytics to open the Analytics tab of the Profiles page, which includes the following functions:

- **Search:** Search across the list of objects.
- **Create:** Opens the New Analytics Profile popup.
- **Edit:** Opens the Edit Analytics Profile popup.
- **Delete:** A profile may only be deleted if it is not currently assigned to a virtual service. An error message will indicate the virtual service referencing the profile. The default system profiles can be edited, but not deleted.

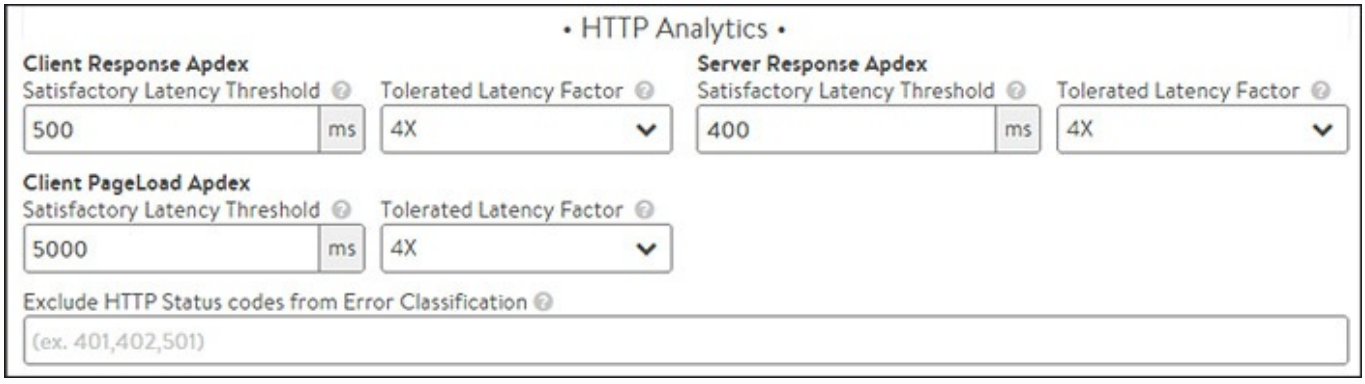
Create an Analytics Profile

The New Analytics Profile and Edit Analytics Profile popups share the same interface.

To create or edit an Analytics Profile:

- **Name:** Enter a unique name for the Analytics Profile.

- **HTTP Analytics:** These settings will only be applied to virtual services configured with an HTTP application profile. For non-HTTP virtual services, these HTTP settings will have no effect.



  Apdex: The concept of Apdex is used extensively by Avi Vantage to capture a client's experience when accessing a virtual service. Apdex is an industry standard for rating a user's experience, which it classifies as Satisfied, Tolerated, or Frustrated.

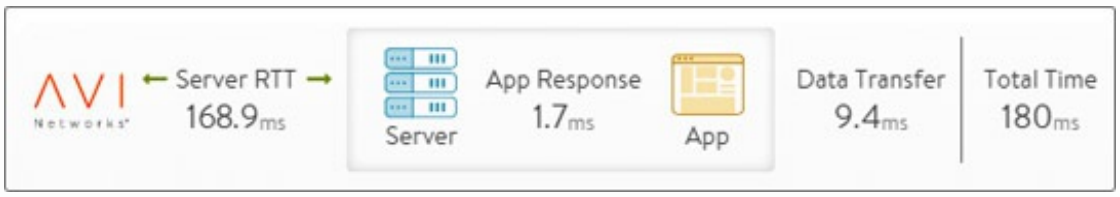
0 - 500 ms	501 - 2000 ms	> 2000 ms
Satisfied	Tolerated	Frustrated

The results are used as part of the performance metric of the pool and virtual service health score. The greater the number of Satisfied responses, the higher the score. See www.apdex.org.

- **Client Satisfactory Latency Threshold:** A client must receive a completed response to an HTTP request within this time frame to be considered satisfied with the transaction time. Using the default of 500ms, any response that is complete with 0-500ms is considered satisfactory. A response time is measured via the End to End Timing's Total Time metric, which includes Client RTT, Server RTT, App Response, and the Data Transfer metrics.

- **Client Tolerated Latency Factor:** The Satisfied Latency Threshold is multiplied by the Tolerated Factor to determine the Tolerated Threshold. If Satisfied is 500ms and the Tolerated Factor is 4x, then 0-500ms is Satisfied, while 500ms to 2000ms is considered a Tolerated response time and anything over 2000ms is considered Frustrated.

- **Server Satisfactory Latency Threshold:** This is the same as the Client Satisfactory Latency Threshold; however, this metric takes the client's latency out of the picture by only measuring Server RTT, App Response, and the Data Transfer time between the server and the Service Engine. This is similar to viewing the End to End Timing of a pool rather than a virtual service. This metric helps differentiate between poor response times due to slow servers versus slow clients.



- **Server Tolerated Latency Factor:** Similar to the Client Tolerated Latency Factor, this option is multiplies the Server Satisfactory Latency Threshold to determine Tolerated responses from servers.

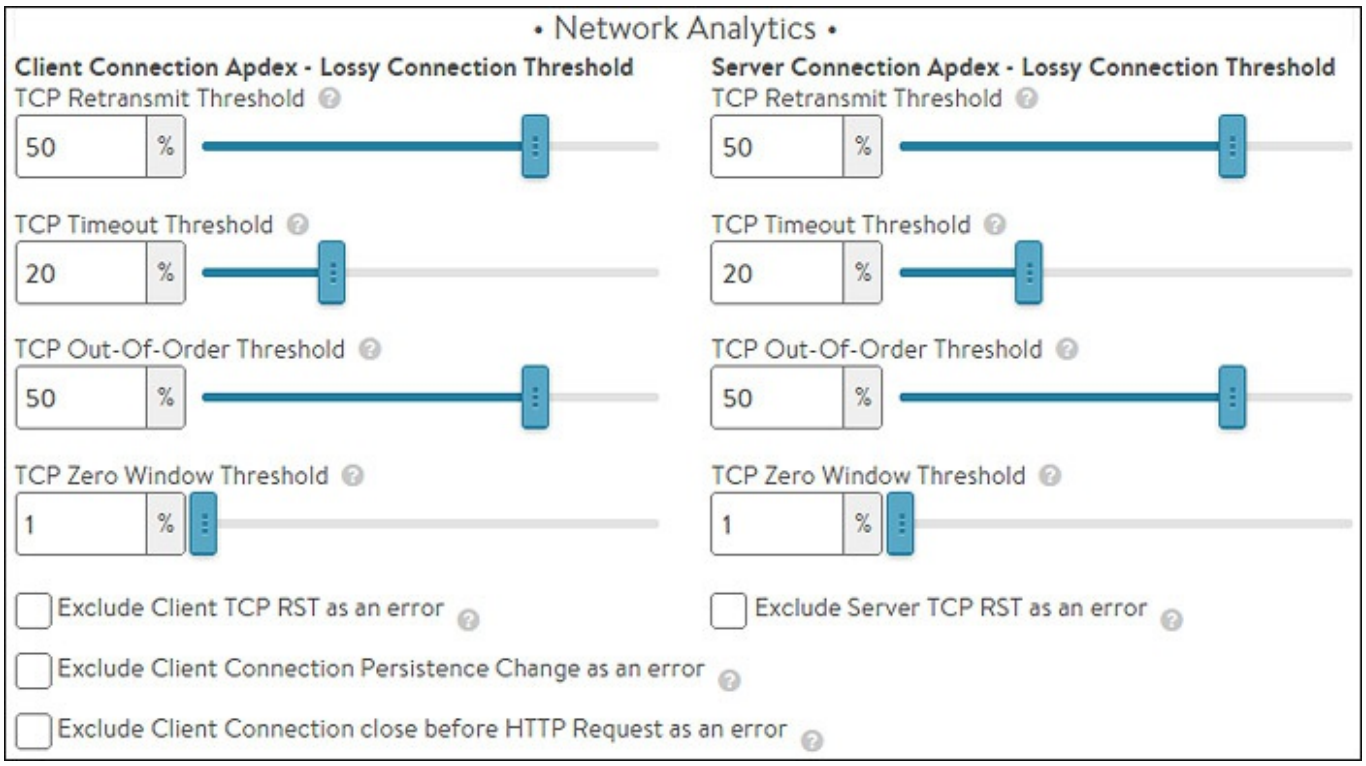
- **Client PageLoad Satisfactory Latency Threshold:** Similar to the Client Satisfactory Latency Threshold, this metric looks at PageLoad times rather than a single HTTP request. PageLoad requires the HTTP virtual service to have the Analytics Type set to Active, which will insert JavaScript into a sampling of HTTP responses. PageLoad

measures the time it takes for a client to download an entire web page, which may include many objects. It also includes the time for DNS lookups, TCP connection setup, object download, and page rendering. For example, Vantage may see satisfied file transfers for objects it is serving, even though clients are seeing errors due to third-party HTML files being slow or having JavaScript errors. This metric is intended catch these issues and incorporate them into the health score.

This metric will catch these issues.

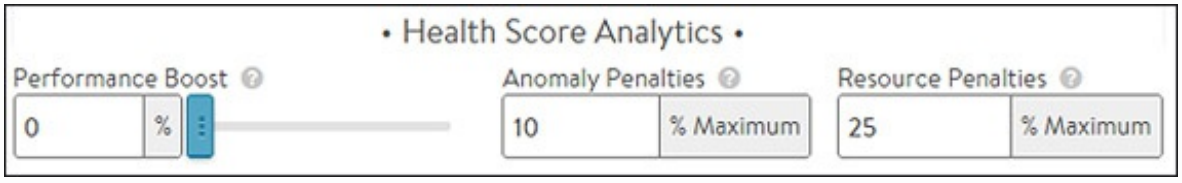
- **Client PageLoad Tolerated Latency Factor:** Similar to the Client Tolerated Latency Factor, this field is used to determine clients using PageLoad that are having a Tolerated or Frustrated experience.
- **Exclude HTTP Status Codes from Error Classification:** By default any 4xx and 5xx responses are considered errors. The greater the percentage of errors, the more the Performance Health Score metric is lowered. Errors are also logged via the Significant client logs. Some of these errors may need to be excluded for certain applications. For instance, Sharepoint will send a 401 error to clients, asking them to first authenticate before accessing the web site.

- **Network Analytics:** Interruptions to TCP connections may happen for a number of reasons. These connections are deemed lossy and are logged via the Significant client logs. They may also reduce the Performance Health Score metric. These metrics are applicable for any virtual service using TCP in proxy mode. Lossy connections may happen on the client or server side of Avi Vantage. Their influence on Health Scores may be adjusted below.



- **TCP Retransmit Threshold:** The TCP connection is considered lossy when more than this percent of packets are retransmitted.
- **TCP Timeout Threshold:** Similar to the previous metric, this option specifically evaluates the number of retransmissions that were required due to timeouts.
- **TCP Out-of-Order Threshold:** The connection is deemed lossy when more than this percentage of packets received from the client are out of order.
- **TCP Zero Window Threshold:** The connection is deemed lossy when greater than this percentage of packets could not be transmitted because the TCP connection window had reduced to zero.

- **Exclude Network Errors:** Some errors may not be abnormal for a given environment. Excluding this issues from the list of errors ensures they will not degrade health score or generate logs.
  - **Client TCP RST:** A graceful TCP shutdown occurs via a FIN/ACK process. Avi Vantage records RST packets as an error unless otherwise excluded.
  - **Client Connection Persistence Change:** Connection persistence change is typically due to a server going offline, forcing Avi Vantage to rebalance connections to new servers. Selecting this option excludes this scenario from the list of errors.
  - **Client Connection Close before HTTP Request:** If the client closes the connection prior to completing an HTTP request, Avi Vantage will record this as an error. Selecting this option excludes this scenario from the list of errors.
  - **Server TCP RST:** A graceful TCP shutdown normally occurs via a FIN/ACK process. Avi Vantage records RST packets as an error unless otherwise excluded. Applications such as Microsoft Exchange may use RST to close connections. Selecting this option will omit server RSTs from the list of errors.
- **Health Score Analytics:** Health scores are assigned to servers, pools, virtual services, and Service Engines. The following settings specifically apply to modifying the health scores of virtual services, whose scores are comprised of performance, anomaly, and resource penalties.



  - **Performance Boost:** Some applications may simply not be able to consistently meet a

response Apdex threshold. For example, an application that relies on a backend database may normally respond within 50ms, but occasional DB queries may take 2 seconds. Rather than set the Client and Server Response Apdex thresholds greater than 2 seconds, instead the Performance metric can be artificially inflated by a small percentage. This allows the Satisfied threshold to remain aggressive while still allowing for occasional slow responses.

- **Anomaly Penalties:** This setting controls how many points Anomalies will deduct from the Performance Score. Anomalies represent risk to the application via inconsistent behavior of clients, traffic volume, or server responses. Lowering the Anomaly Penalty places greater emphasis on the Performance Score and Resource Penalties.
- **Resource Penalties:** Resources that are constrained will increase the Resource Penalty score. Examples include the CPU, memory, or disk utilization of a Service Engine or a server running on a virtual machine.

Groups

Avi Vantage uses “groups” to separate long lists from the complexity of the feature they may be applied against. This is particularly useful for simplifying the automation of API calls. This article describes the following types of groups:

- IP group: named set of IP addresses
- String group: named set of string values
- MicroService group: named set of microservices

You may change the contents of these types of groups at any time while the system is running without interrupting existing traffic. Objects that reference the group will use the updated version for subsequent queries they perform against it.

IP Group

IP groups are comma-separated lists of IP addresses that may be referenced by profiles, policies, and logs. IP groups may contain the following types of entries:

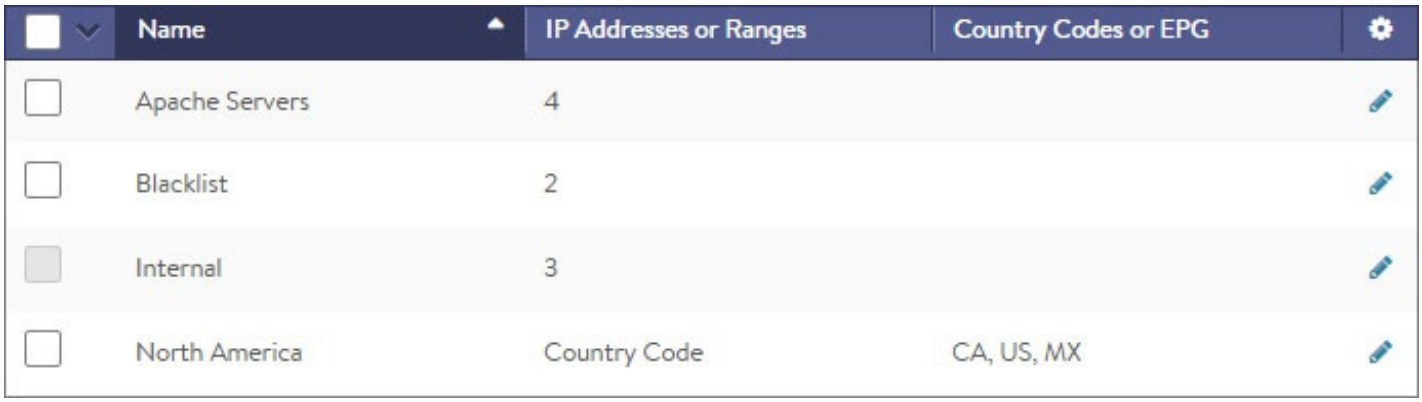
- IPv4 address
- IP range
- IP/Mask

- Country code
- End-point group

The following are examples of IP groups used elsewhere within the Vantage configuration:

- Network security policy: may be configured to drop any clients coming from a blacklist of IP addresses. Rather than maintain a long list within the policy, Vantage maintains the rule logic of that policy separately from the list of addresses, which are kept in the IP group. A user can be granted a role that allows them to update the list of IP addresses without being able to change the policy itself.
- Logs: classify clients by their IP address and match them against an included geographic country location database. You may override this database by using an IP group to create specific mappings such as internal IP addresses. For instance, LA_Office may contain 10.1.0.0/16, while NY_Office contains 10.2.0.0/16. Logs will show these clients as originating from these locations. Logs searches may also be performed on the group name such as LA_Office.

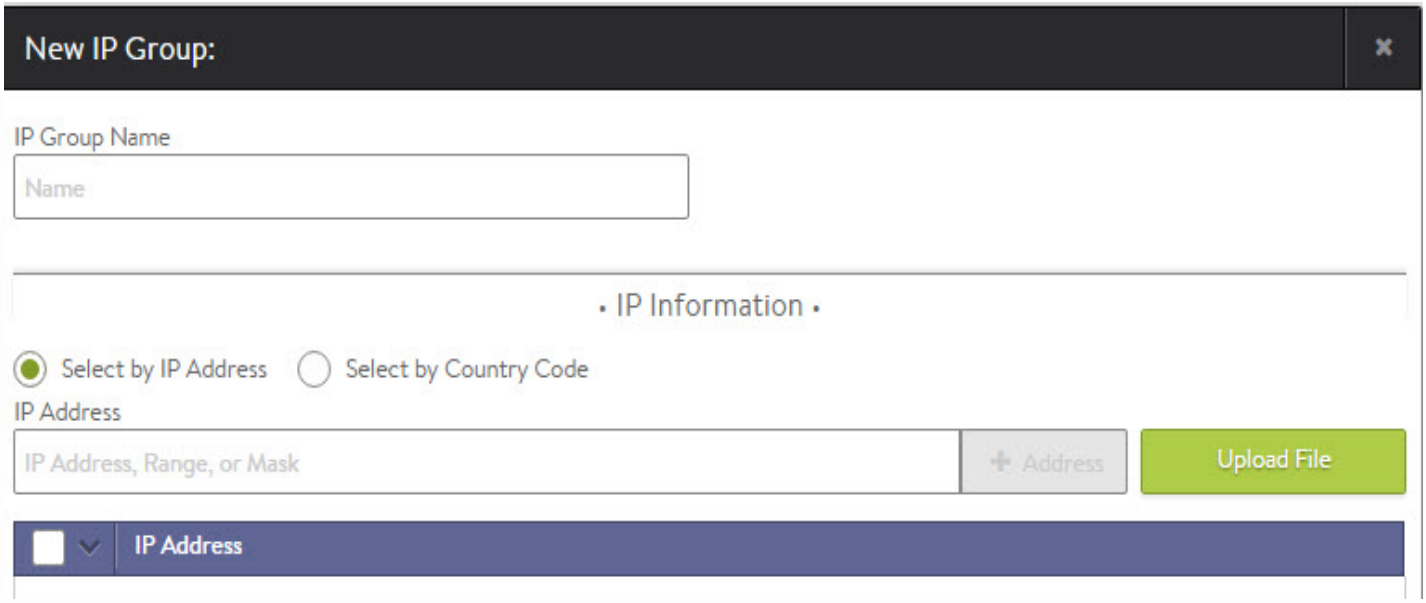
The table on this tab contains the following information for each IP group:



- **Name:** Name of the IP group.
- **IP Address or Ranges:** Number of IP address, networks, or address ranges.
- **Country Code or EPG:** Any configured country codes or end point groups are listed.

Create an IP Group

The Create IP Group and Edit IP Group popups share the same interface.



To create or edit an IP group:

- **Name:** Enter a unique name for the IP group.
- **Select by IP Address:** Add one or more addresses to the list. Any combination of the following methods may be used:
  - **IP Address:** Comma separated list: 10.0.0.1, 10.0.0.5
  - **Range:** 10.0.0.10–10.0.0.20
  - **Mask:** 10.1.0.0/16
  - **Upload File:** Click the Upload File button on the right to navigate to and upload a comma separated value (CSV) file that contains any combination of IP addresses, range, and/or masks.
- **Select by Country Code or EPG:** Select one or more countries, or type the country name into the search field to filter. Countries may not be combined within an IP group with individual IP addresses. An IP group that contains countries may also not be used as a list of servers for pool membership.For cloud environments set up for Cisco ACI, end point groups are shown. EPGs are IP addresses or groups configured within the Cisco APIC controller and automatically imported into Vantage.

String Group

String groups contain arbitrary lists of comma-separated text that may consist of HTTP URIs, names, or other content to be filtered against for many purposes such as black or white lists.

An HTTP security policy may use a string group to filter allowed HTTP URIs.

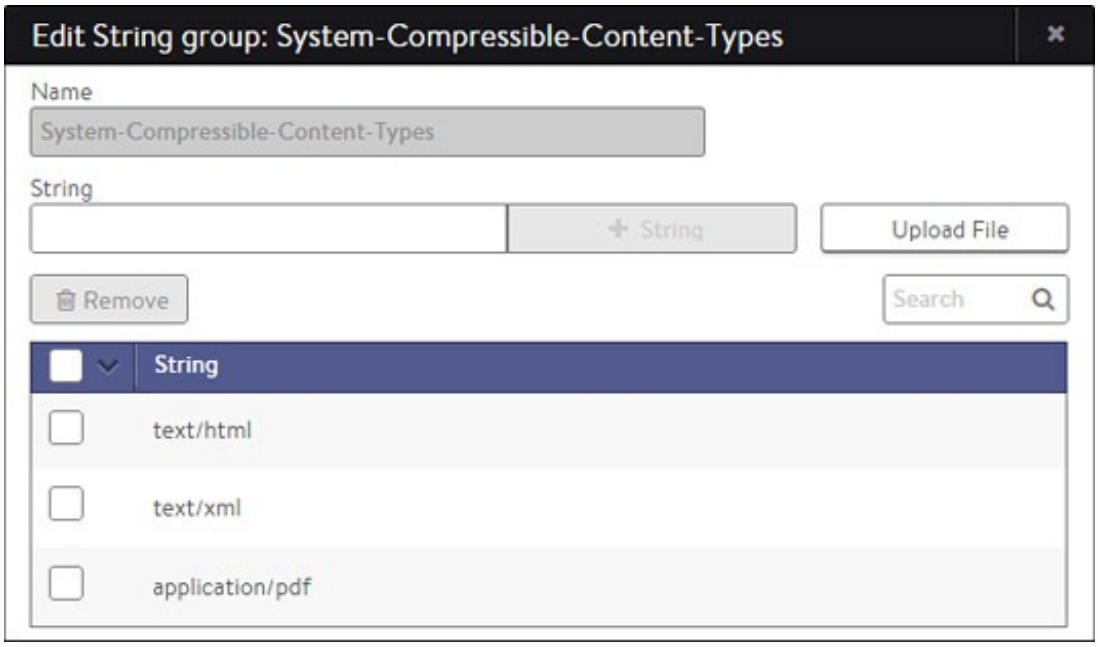
An HTTP application profile may use a string group to identify mobile users to ensure they receive more aggressively compressed HTML files.

The table on this tab contains the following information for each string group:

- **Name:** Name of the string group.
- **Number of Strings:** Number of strings within a string group.

Create a String Group

The Create String Group and Edit String Group popups share the same interface.



To create or edit a string group:

- **Name:** Enter a unique name for the String Group in the Name field.
- **Strings:** Add one or more text strings to the list. You may use any combination of the following methods:
  - **String:** Enter any text in the field, and then click the + String button to add.
  - **Upload File:** Click the Upload File button on the right to navigate to and upload a comma separated value (CSV) file that contains text.

MicroService Group

This article will be updated with more information on this feature.

Security

SSL Certificates

Avi Vantage supports terminating client SSL and TLS connections at the virtual service. This requires Vantage to send a certificate to clients that authenticates the site and establishes secure communications. A virtual service that handles secure connections will require both of the following:

- **SSL/TLS profile** - determines the supported ciphers and versions.
- **SSL certificate** - presented to clients connecting to the site. SSL certificates may also be used to present to administrators connecting to the Vantage web interface or API, and also for the Avi Service Engine (SE) to present to servers when SE-to-server encryption is required.

The SSL/TLS Certificates page allows import, export, and generation of new SSL certificates or certificate requests. Newly-created certificates may be either self-signed by Avi Vantage or created as a Certificate Signing Request (CSR) that must be sent to a trusted Certificate Authority (CA), who then generates a trusted certificate.

Creating a self-signed certificate generates both the certificate and a corresponding private key.

Imported existing certificates are not valid until a matching key has been supplied.

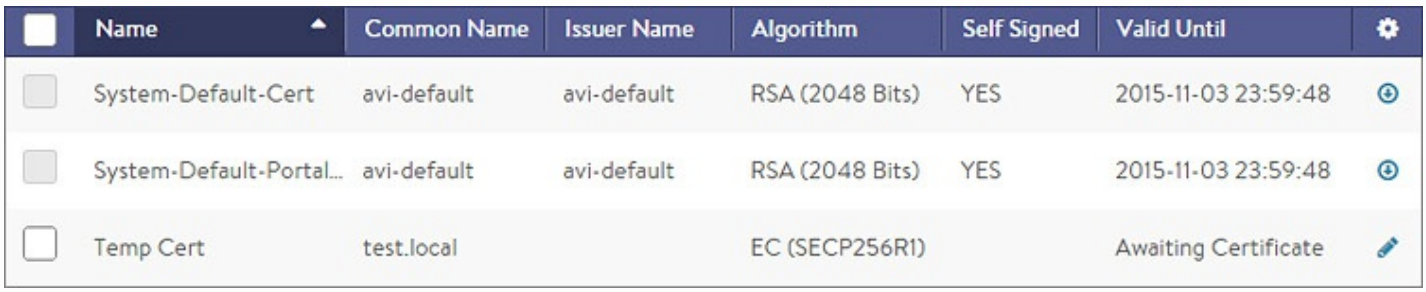
Avi Vantage supports PEM and PKCS12 formats for certificates.

SSL/TLS Certificates Page

Select Templates > SSL/TLS Certificates to open the SSL/TLS Certificates page. This tab includes the following functions:

- **Search:** Search across the list of objects.
- **Create:** Opens the Create Certificate popup.
- **Edit:** Opens the Edit Certificate popup. Only incomplete certificates that do not have a corresponding key are able to be edited.
- **Export:** The down arrow icon exports a certificate and corresponding private key.
- **Delete:** A certificate may only be deleted if it is not currently assigned to a virtual service. An error message will indicate the virtual service referencing the certificate.

The table on this tab contains the following information for each certificate:

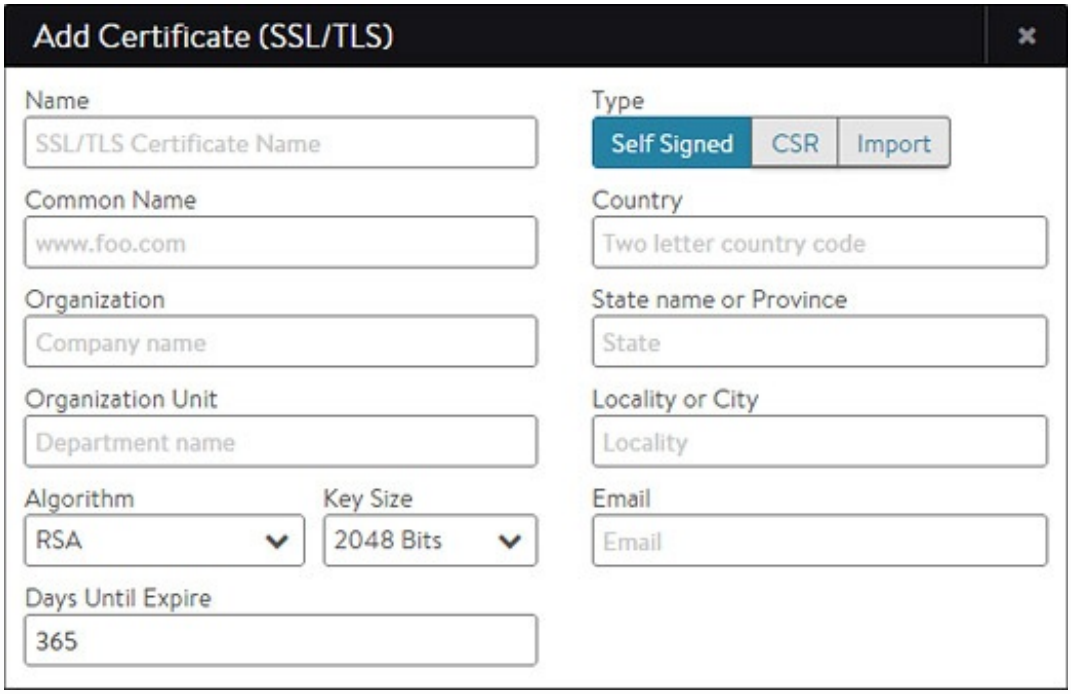


Name	Common Name	Issuer Name	Algorithm	Self Signed	Valid Until
System-Default-Cert	avi-default	avi-default	RSA (2048 Bits)	YES	2015-11-03 23:59:48
System-Default-Portal...	avi-default	avi-default	RSA (2048 Bits)	YES	2015-11-03 23:59:48
Temp Cert	test.local		EC (SECP256R1)		Awaiting Certificate

- **Name:** User-friendly name of the certificate. Mouse over the name of the cert will display any intermediate cert that has been automatically associated with the cert.
- **Status**: The known status of the certificate. This is green for good, or yellow, orange, or red if the cert is expiring soon or has already expired, and gray if the cert is incomplete.
- **Common Name:** Fully qualified name of the site to which the certificate applies. This entry must match the hostname the client will enter in their browser in order for the site to be considered trusted.
- **Issuer Name:** Name of the Certificate Authority.
- **Algorithm:** This will be either EC (Elliptic Curve) or RSA.
- **Self Signed:** Whether the certificate is self-signed by Vantage or signed by a Certificate Authority.
- **Valid Until:** Date and time when the certificate expires.

Create Certificate

Click New in the SSL/TLS Certificates page to open the Add Certificate (SSL/TLS) popup.



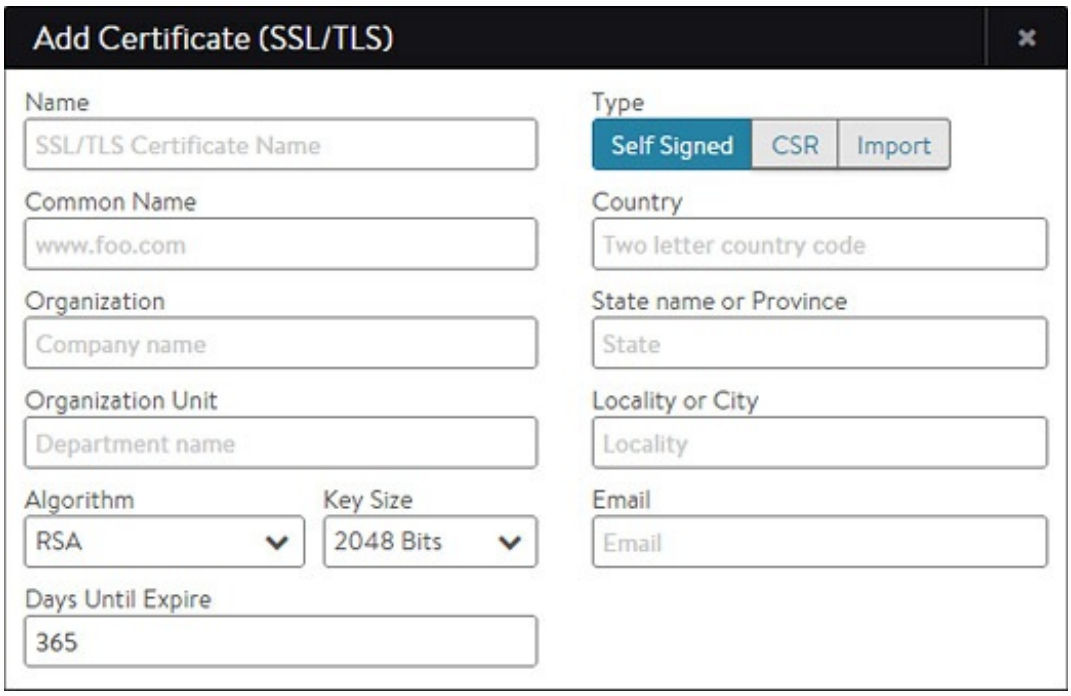
To create a new certificate:

- **Name:** Enter a unique, user-friendly name for the certificate in the Name field.
- **Common Name:** Enter the fully qualified name of the site, such as www.avinetworks.com. This entry must match the hostname the client entered in the browser in order for the site to be considered trusted.
- **Type:** Select the type of certificate to create:
    - **Self Signed:** Quickly create a test certificate that is signed by Vantage. Client browsers will display an error that the certificate is not trusted. If the HTTP application profile has HTTP Strict Transport Security (HSTS) enabled, clients will not be able to access a site with a self signed certificate.
    - **CSR:** Create a valid certificate by first creating the certificate request. This request must be sent to a certificate authority, which will send back a valid certificate that must be imported back into Vantage.
    - **Import:** Import a completed certificate that was either received from a certificate authority or exported from another server.
- Input all of the required information required for the type of certificate you are creating:
    - Self-Signed Certificates
    - CSR Certificates
    - Importing Certificates

Self-Signed Certificates

Avi Vantage can generated self-signed certificates. Client browsers do not trust these certificates

and will warn the user that the Virtual Service's certificate is not part of a trust chain. Self-signed certificates are good for testing or environments where administrators control the clients and can safely bypass the browser's security alerts. Public websites should never use self-signed certificates.



If you selected **Self-Signed** as the certificate Type in the Add Certificate popup, then:

- Enter the following information:
  - **Organization:** Company or entity registering the certificate, such as Avi Networks, Inc. (optional).
  - **Organization Unit:** Group within the organization that is responsible for the certificate, such as Development (optional).
  - **Country:** Country in which the organization is located (optional).
  - **State:** State in which the organization is located (optional).
  - **Locality:** City of the organization (optional).
  - **Email:** The email contact for the certificate (optional).
  - **Algorithm:** Select either EC (Elliptic Curve) or RSA. RSA is older and considered less secure than EC, but is more compatible with a broader array of older browsers. EC is newer, less computationally expensive, and generally more secure; however, it is not yet accepted by all clients. Avi Vantage allows a Virtual Service to be configured with two certificates at a time, one each of RSA and EC. This allows Avi Vantage to negotiate the

optimal algorithm with the client. If the client supports EC, then Avi Vantage will prefer this algorithm, which gives the added benefit of natively supporting Perfect Forward Secrecy for better security.

Key Size: Select the level of encryption to be used for handshakes, as follows:

- 2048 Bit is recommended for RSA certificates.
- SECP256R1 is recommended for EC certificates.

Higher values may provide better encryption but increase the CPU resources required by both Avi Vantage and the client.

- Finish entering the required information in the Add Certificate popup.

CSR Certificates

The Certificate Signing Request (CSR) is the first of three steps involved in creating a valid SSL/TLS certificate. The Request contains the same parameters as a Self-Signed Certificate; however, Avi Vantage does not sign the completed certificate. Rather, it must be signed by a Certificate Authority that is trusted by client browsers.

If you selected CSR as the certificate Type in the Add Certificate popup, then:

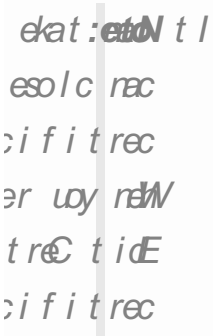
- Enter the following information:
  - **Organization:** Company or entity registering the certificate, such as Avi Networks
  - **Organization Unit:** Group within the organization that is responsible for the certificate, such as Development.
  - **Country:** Country in which the organization is located.
  - **State:** State in which the organization is located.
  - **Locality:** City of the organization.
  - **Email:** The email contact for the certificate.
  - **Algorithm:** Select either EC (Elliptic Curve) or RSA. RSA is older and considered less secure than EC, but is more compatible with a broader array of older browsers. EC is newer, less computationally expensive, and generally more secure; however, it is not yet accepted by all clients. Avi Vantage allows a Virtual Service to be configured with two certificates at a time, one each of RSA and EC. This allows Avi Vantage to negotiate the optimal algorithm with the client. If the client supports EC, then Avi Vantage will preference this algorithm, which gives the added benefit of natively supporting Perfect

Forward Secrecy for better security.

- **Key Size:** Select the level of encryption to be used for handshakes, as follows:
  - 2048 Bit is recommended for RSA certificates.
  - SECP256R1 is recommended for EC certificates.

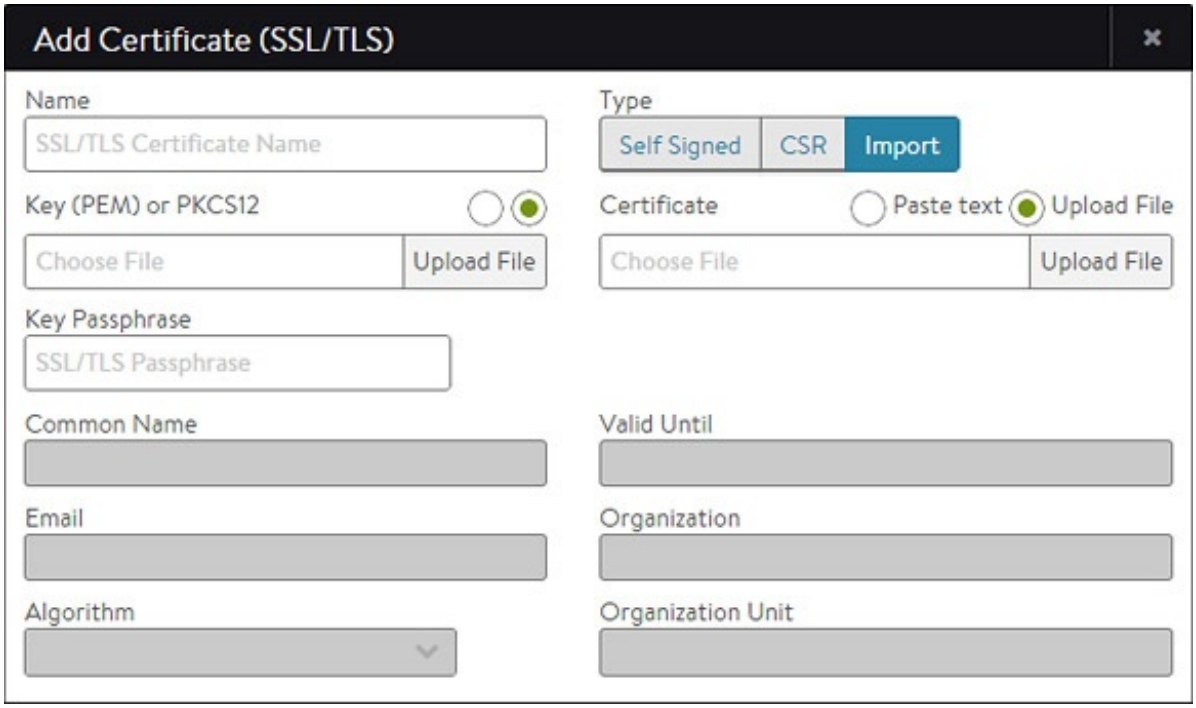
  Higher values provide better encryption but increase the CPU resources required by both Avi Vantage and the client.

- Finish entering the required information in the Add Certificate popup, and then click Generate to save and generate the CSR.
- Forward the completed CSR to any trusted Certificate Authority (CA), such as Thawte or Verisign, by selecting the Certificate Signing Request at the bottom left of the Add Certificate popup and then either copying and pasting it directly to the CA's website or saving it to a file for later use.
- Once the CA issues the completed certificate, you may either pate or upload it into the Certificate field at the bottom right of the Add Certificate popup.



Import Certificates

You may directly import an existing PEM or PKCS12 SSL/TLS certificate into Avi Vantage (such as from another server or load balancer). A certificate will have a corresponding private key, which must also be imported.



If you selected Import as the certificate type in the Add Certificate popup, then:

- **Key:** Add the private key via one of the two methods below. Switch between the two methods via the green radio buttons.
  - **Upload File:** Click the Upload File button, select the PEM or PKCS12 file, then click the green Validate button to parse the file. If the upload is successful, the Key field will be populated.
  - **Paste:** Copy and paste a PEM key into the Key field. Be careful to not introduce extra characters in the text, which can occur when using some email clients or rich text editors. If you copy and paste the key and certificate together as one file, click the Validate button to parse the text and populate the Certificate field.

    PKCS12 is a binary format, which cannot be copied/pasted or use this method.

- **Certificate:** Add the cert in the Certificate field if this is not already populated from the previous step. You may do this by copying/pasting or file upload, as described above.
- **Key Passphrase:** If desired, you may add and validate a Key Passphrase to encrypt the private key.

- **Import:** Select Import to finish adding the new certificate and key. The key will be embedded with the certificate and treated as one object within the Avi Vantage UI.

Certificate Authority

Certificates require a trusted chain of authority to be considered valid. If the certificate used is directly generated by a certificate authority that is known to all client browsers, no cert chain is required. However, if there are multiple levels required, an intermediate certificate may be necessary. Clients will often traverse the path indicated by the certificate to validate on their own if no chain cert is presented by a site, but this adds additional DNS lookups and time for the initial site load. The ideal scenario is to present the chain certs along with the site cert.

If a chain certificate, or rather a certificate for a certificate authority, is uploaded via the certificate > import in the certificates page, it will be added to the Certificate Authority section. Vantage will automatically build the cert chain if it detects a next link in the chain exists.

To validate a certificate has been attached to a chain cert, mouse over the certificate's name in the SSL Certificates table at the top of the page.

Vantage supports multiple chain paths. For instance, an RSA cert and an EC cert are both named www.avinetworks.com. Each may share the same CA issuer, or they may be chained to different issuers.

SSL Profile

Avi Vantage supports the ability to terminate SSL connections between the client and the Virtual Service, and to enable encryption between Avi Vantage and the back-end servers. The SSL/TLS Profile contains the list of accepted SSL versions and the prioritized list of SSL ciphers. Both an SSL/TLS Profile and an SSL Certificate must be assigned to the Virtual Service when configuring it to terminate client SSL/TLS connections. If you will also be encrypting traffic between Avi Vantage and the servers, an SSL/TLS Profile must be assigned to the Pool. When creating a new Virtual Service via the Basic mode, the default system SSL/TLS profile is automatically used.

SSL termination may be done on any service port; however, browsers will assume the default port of 443. The best practice is to configure a Virtual Service to accept both HTTP and HTTPS by creating a service on port 80, selecting the + icon to add an additional service port, and then setting the new service port to 443 with SSL enabled. A redirect from HTTP to HTTPS is generally preferable, which can be done via a Policy or by using the System-HTTP-Secure application profile.

Each SSL/TLS profile contains default groupings of supported SSL ciphers and versions that may be used with RSA or an Elliptic Curve certificate, or both. Ensure that any new SSL/TLS profile you create includes ciphers that are appropriate for the certificate type that will be used. The

default SSL/TLS profiles included with Vantage provide a broad range of security; for example, the Standard Profile will work for typical deployments.

Creating a new SSL/TLS profile or using an existing Profile entails various trade-offs between security, compatibility, and computational expense. For example, increasing the list of accepted ciphers and SSL versions increases the compatibility with clients while also potentially lowering security.

SSL Profile Settings

Select Templates > Profiles > SSL/TLS to open the SSL/TLS Profiles tab. This tab includes the following functions:

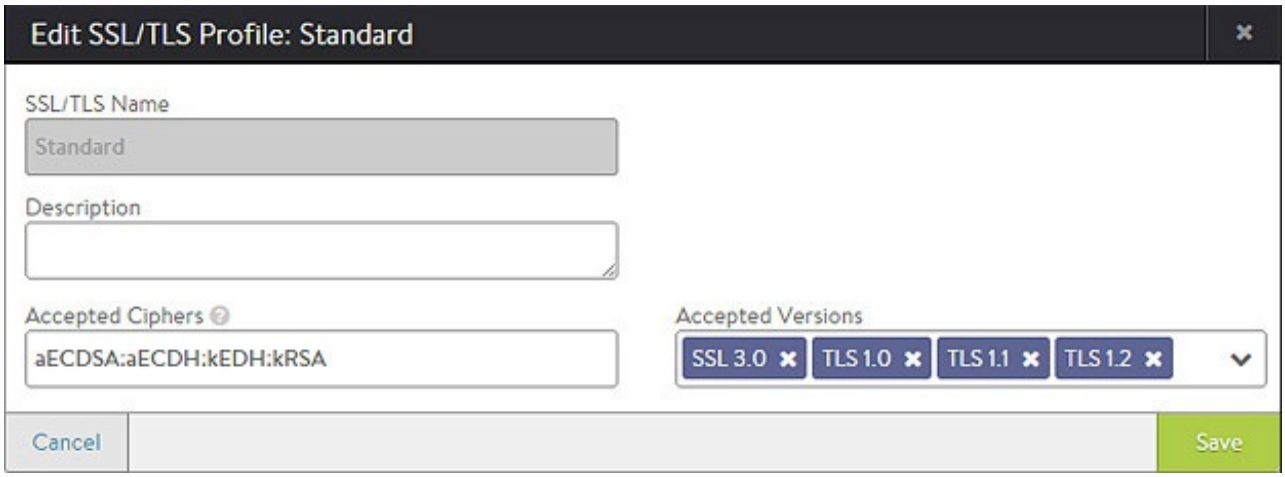
- **Search:** Search across the list of objects.
- **Create:** Opens the New SSL/TLS Profile popup.
- **Edit:** Opens the Edit SSL/TLS Profile popup.
- **Delete:** An SSL/TLS profile may only be deleted if it is not currently assigned to a virtual service. An error message will indicate the virtual service referencing the profile. The default system profiles can be modified, but not deleted.

The table on this tab provides the following information for each SSL/TLS profile:

- **Name:** Name of the profile.
- **Accepted Ciphers:** List of ciphers accepted by the profile, including the prioritized order.
- **Accepted Versions:** SSL and TLS versions accepted by the profile.

Create an SSL Profile

To create or edit an SSL profile:



- **Name:** Enter a unique name for the SSL/TLS Profile in the Name field.
- **Accepted Ciphers:** Enter the list of accepted ciphers in the Accepted Ciphers field. Each cipher entered must conform to the cipher suite names listed at OpenSSL. Separate each cipher with a colon. For example, AES:3DES means that this Profile will accept the AES and 3DES ciphers. When negotiating ciphers with the client, Avi Vantage will preference ciphers in the order listed. You may use an SSL/TLS profile with both an RSA and an Elliptic Curve certificate. These two types of certificates can use different types of ciphers, so it is important to incorporate ciphers for both types. Selecting only the most secure ciphers may incur higher CPU load on Vantage and may also reduce compatibility with older browsers.
- The **Accepted Versions** pull-down menu allows you to select one or more SSL/TLS versions to add to this profile. Chronologically, TLS v1.0 is the oldest supported, and TLS v1.2 is the newest. SSL v3.0 is no longer support as of Avi Vantage v15.2. In general with SSL, older versions have many known vulnerabilities while newer versions have many undiscovered vulnerabilities. As with any security, Avi recommends diligence to understand the dynamic nature of security and to ensure that Avi Vantage is always up to date. Some SSL ciphers are dependent on specific versions of SSL or TLS supported. See OpenSSL for more information about this subject.

PKI Profile

The Public Key Infrastructure (PKI) profile allows configuration of Certificate Revocation List (CRLs), and the process for updating the lists. The PKI profile may be used to validate clients and server certificates.

- **Client Certificate Validation:** Vantage supports the ability to validate client access to an HTTPS site via client SSL certificates. Clients will present their certificate when accessing a virtual service, which will be matched against a CRL. If the certificate is valid and the clients are not on the list of revoked certificates, they will be allowed access to the HTTPS site.

  Client certificate validation is enabled via the HTTP profile's Authentication tab. The HTTP profile will reference the PKI profile for specifics on the Certificate Authority (CA) and the CRL. A PKI profile may be referenced by multiple HTTP profiles.
- **Server Certificate Validation:** Similar to validating a client certificate, Vantage can validate the certificate presented by a server, such as when an HTTPS health monitor is sent to a server. Server certificate validation uses the same PKI profile to validate the certificate presented.

  Server certificate validation can be configured by enabling SSL within the desired pool, then specifying the PKI Profile.

PKI Profile Settings

Select Templates > Security > PKI Profile to open the PKI tab. This tab includes the following functions:

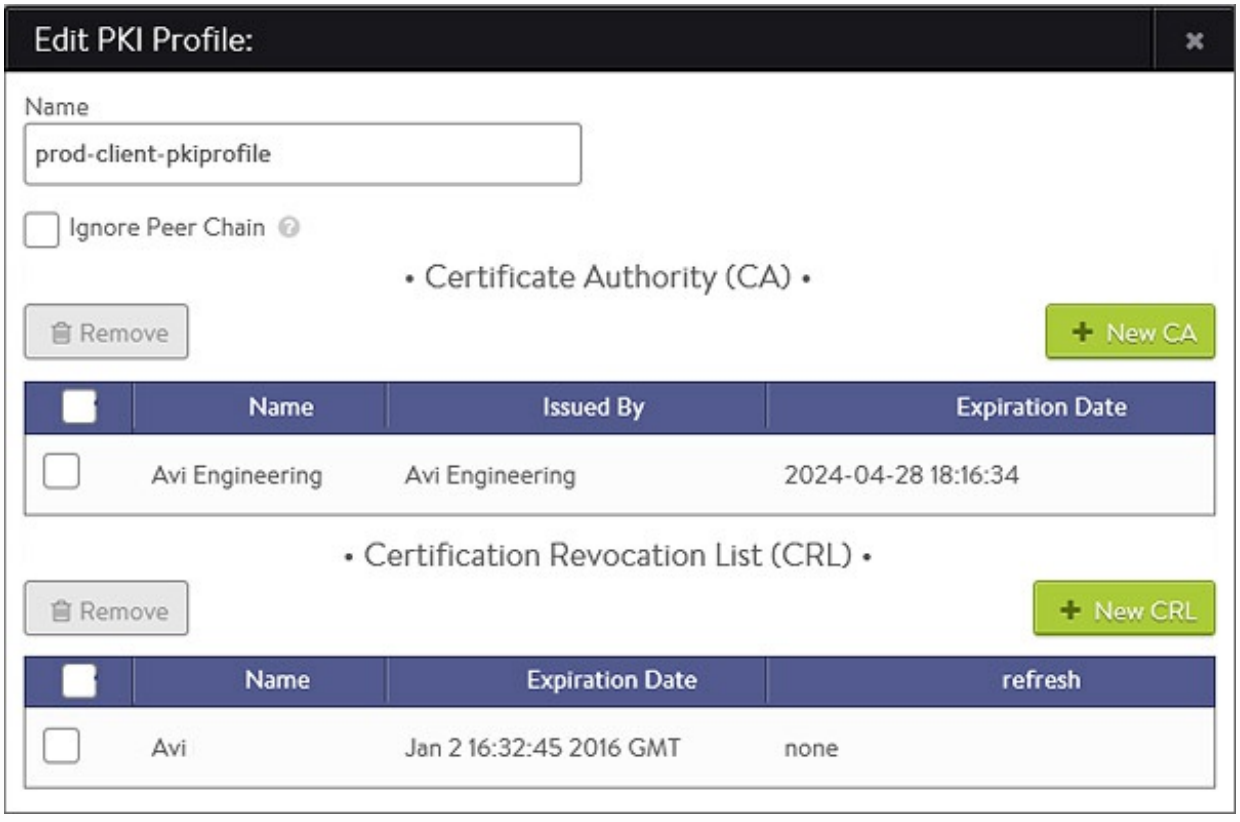
- **Search:** Search across the list of objects.
- **Create:** Opens the New PKI Profile popup.
- **Edit:** Opens the Edit PKI Profile popup.
- **Delete:** A PKI profile may only be deleted if it is not currently assigned to an HTTP profile. An error message will indicate the HTTP profile referencing the PKI profile.

The table on this tab provides the following information for each PKI Profile:

- **Name:** Name of the Profile.
- **Certificate Authority:** Denotes if a CA has been attached to the PKI Profile.
- **Certificate Revocation List:** Revocation lists (CRLs) that have been attached to the PKI Profile.

Create a PKI Profile

To create or edit a PKI Profile:



- **Name:** Enter a unique name for the PKI profile.
- **Ignore Peer Chain:** When set to true, the certificate validation will ignore any intermediate certificates that might be presented. The presented certificate is only checked against the final root certificate for revocation.When this option is disabled (default), the certificate must present a full chain which is traversed and validated, starting from the client or server presented cert to the terminal root cert. Each intermediate cert must be validated and matched against a CA cert included in the PKI profile.
- **Certificate Authority:** Add a new certificate from a trusted Certificate Authority. If more than one CA are included in the PKI profile, then a client's certificate needs only to match to one of them to be valid. A client's certificate must match the CA as the root of the chain. If the presented cert has an intermediate chain, then each link in the chain must be included here. See Ignore Peer Chain to ignore intermediate validation checking.
- **Client Revocation List:** The CRL allows invalidation of certificates, or more specifically the certificate's serial number. The revocation list may be updated by manually uploading a new CRL, or by periodically downloading from a CRL server. If a client or server certificate is found to be in the CRL, the SSL handshake will fail, with a resulting log created to provide further information about the handshake.
    - **Server URL:** Specify a server from which to download CRL updates. Access to this server will be done from the Controller IP addresses, which means they will require firewall access to this destination. The server may be an IP address, or an FQDN along with an HTTP path, such as www.avinetworks.com/crl.
    - **Refresh Time:** After the elapsed period of time, Avi will automatically download an updated version of the CRL. If no time is specified, Avi will download a new CRL at the current CRL's lifetime expiration.
    - **Upload CRL File:** Manually upload a CRL. Subsequent CRL updates can be done by manually uploading newer lists, or configuring the Server URL and Refresh Time to automate the process.

Authentication Profile

The Authentication profile ("auth profile") allows for configuration of clients into a Virtual Service via HTTP basic authentication. As of version 15.1, Vantage supports remote authentication via LDAP.

The authentication profile is enabled via the HTTP basic authentication setting of a virtual service's Advanced properties tab.

Avi also supports client authentication via SSL client certificates, which is configured the HTTP

Profile's Authentication section.

Auth Profile Settings

Select Templates > Security > Auth Profile to open the Auth tab. This tab includes the following functions:

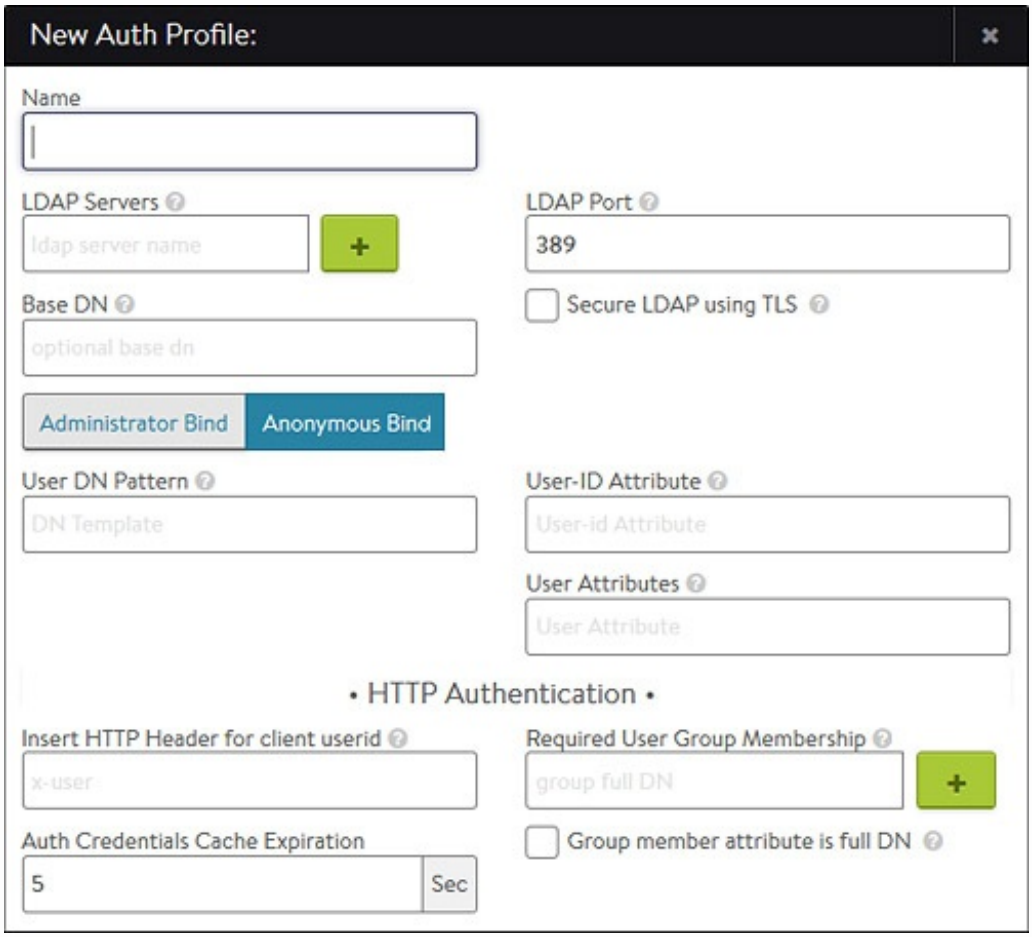
- **Search:** Search across the list of objects.
- **Create:** Opens the Create/Edit window.
- **Edit:** Opens the Create/Edit window.
- **Delete:** An Auth profile may only be deleted if it is not currently assigned to a virtual service or in use by Vantage for administrative authentication.

The table on this tab provides the following information for each auth profile:

- **Name:** Name of the profile.
- **Type:** The Type will be LDAP.

Create an Authentication Profile

To create or edit an authentication profile:



- **Name:** Enter a unique name.

- **LDAP Servers:** Configure one or more LDAP servers by adding their IP addresses.
- **LDAP Port:** The service port to use when communicating with the LDAP servers. This is typically 389 for LDAP or 636 for LDAPS (SSL).
- **Secure LDAP using TLS:** Enable startTLS for secure communications with the LDAP servers. This may require a service port change.
- **Base DN:** LDAP Directory Base Distinguished Name. Used as default for settings where DN is required but was not populated like User or Group Search DN.
- **Anonymous Bind:** Minimal LDAP settings that are required to verify User authentication credentials by binding to LDAP server. This option is useful when you do not have access to administrator account on the LDAP server.
    - **User DN Pattern:** LDAP user DN pattern is used to bind an LDAP user after replacing the user token with real username. The pattern should match the user record path in the LDAP server. For example, cn=,ou=People,dc=myorg,dc=com is a pattern where we expect to find all user records under ou “People”. When searching LDAP for a specific user, we replace the token with username.
    - **User Token:** An LDAP token is replaced with real user name in the user DN pattern. For example, in User DN Pattern is configured as “cn=-user-,ou=People,dc=myorg,dc=com”, the token value should be -user-.
    - **User ID Attribute:** LDAP user ID attribute is the login attribute that uniquely identifies a single user record. The value of this attribute should match the username used at the login prompt.
    - **User Attributes:** LDAP user attributes to fetch on a successful user bind. These attributes are only for debugging.
- **Administrator Bind:** LDAP administrator credentials configured under LDAP Directory Settings below are used to bind Avi as an admin when querying LDAP for Users or Groups.
    - **Admin Bind DN:** Full DN of LDAP administrator. Admin bind DN is used to bind to an LDAP server. Administrators should have sufficient privileges to search for users under user search DN or groups under group search DN.
    - **Admin Bind Password:** Administrator password. Password expiration or change is not handled. The password is hidden from rest API and CLI.
    - **User Search DN:** LDAP user search DN is the root of search for a given user in the LDAP directory. Only user records present in this LDAP directory sub-tree are allowed

for authentication. Base DN value is used if this value is not configured.

- **Group Search DN:** LDAP group search DN is the root of search for a given group in the LDAP directory. Only matching groups present in this LDAP directory sub-tree will be checked for user membership. Base DN value is used if this value is not configured.
- **User Search Scope:** LDAP user search scope defines how deep to search for the user starting from user search DN. The options are search at base, search one-level below or search the entire sub-tree. The default option is to search one-level deep under user search DN.
- **Group Search Scope:** LDAP group search scope defines how deep to search for the group starting from the group search DN. The default value is the entire subtree.
- **User ID Attribute:** LDAP user ID attribute is the login attribute that uniquely identifies a single user record. The value of this attribute should match the username used at the login prompt.
- **Group Member Attribute:** LDAP group attribute that identifies each of the group members. For example, member and memberUid are commonly used attributes.
- **User Attributes:** LDAP user attributes to fetch on a successful user bind. These attributes are only for debugging.

- **Insert HTTP Header for Client UserID:** Insert an HTTP header into the client request before it is sent to the destination server. This field is used to name the header. The value will be the client's User ID. This same UserID value will also be used to populate the User ID field in the Virtual Service's logs.
- **Required User Group Membership:** User should be a member of these groups. Each group is identified by the DN, for example,'cn=testgroup,ou=groups,dc=LDAP,dc=example,dc=com'
- **Auth Credentials Cache Expiration:** The max allowed length of time a client's authentication is cached.
- **Group Member Attribute Is Full DN:** Group member entries contain full DNs instead of just User ID attribute values.

Scripts

Extensive behavior customization and automation can be done by utilizing scripts for both the control plane with the Python-based ControlScripts, and the data plane via Lua-based

DataScripts.

DataScript

DataScripts are a powerful mechanism for customizing the behavior of Avi Vantage on a per virtual service, or even per client, basis. DataScripts are lightweight scripts coded in Lua. These scripts can be executed against each client making a TCP connection, an HTTP request or response, or other events within the dataplane.

One or more DataScripts may be attached to the rules section of a virtual service.

Scripts may be uploaded or copy/pasted into either the Request Event Script or Response Event Script section. For instance, to restrict access to the secure directory, the following text would be pasted into the Request Event Script section.

```
if avi.http.uri == "/secure/" then
   avi.http.send(403)
end
```

See the Avi DataScript Guide for complete documentation of commands and example DataScripts.

ControlScript

ControlScripts are Python-based scripts which execute on the Vantage Controllers. They are initiated by Alert Actions, which themselves are triggered by events within the system. Rather than simply alert an admin that a specific event has occurred, the ControlScript can take specific action, such as altering the Vantage configuration or sending a custom message to an external system, such as telling VMware's vCenter to scale out more servers if the current servers have reached resource capacity and are incurring lowered health scores.

To create or edit a ControlScript, select create and paste the new script into the window or upload the .py file.

Once the ControlScript is uploaded to Vantage, apply the script via Operations > Alert > Alert Action page.

ControlScripts are executed with limited privileges within the Linux subsystem of the Avi Controller. See standard Python documentation for examples and definitions of Python commands. Vantage configuration changes may be made by API calls from Linux to Avi Vantage via standard API mechanisms.

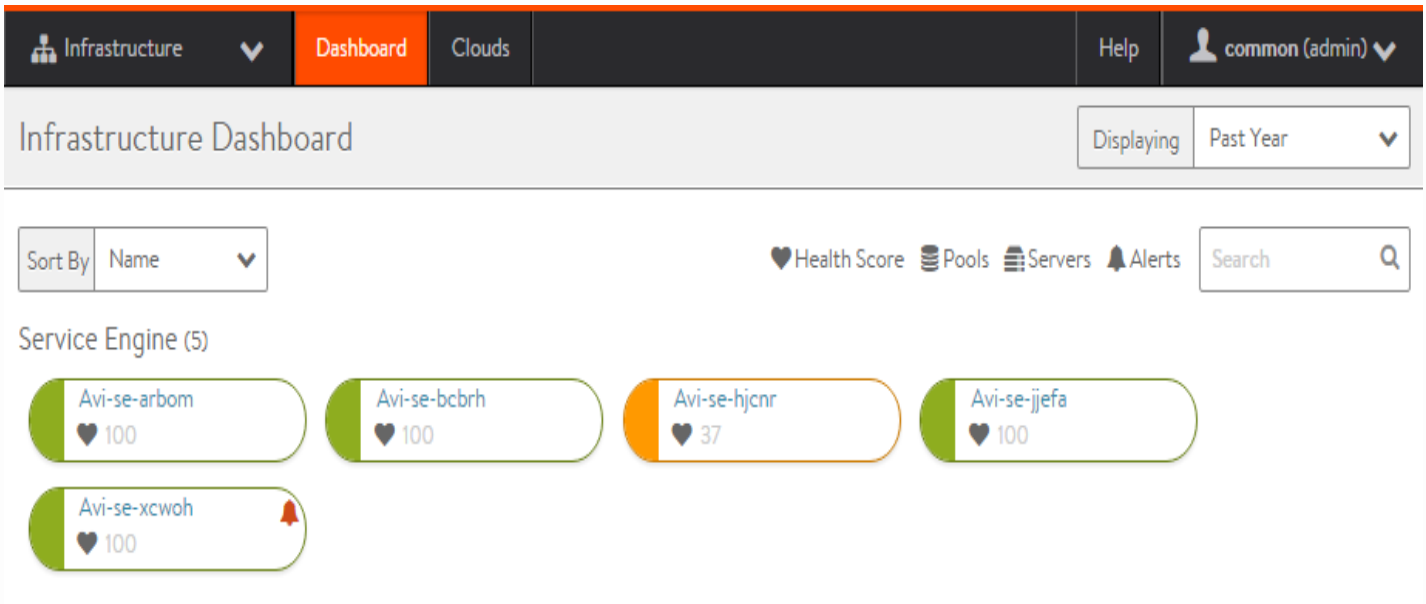
Infrastructure

Avi Vantage may be deployed in multiple environments for the same system. Each environment is called a “cloud”. During the initial configuration of the Controller, a cloud is created by selecting the deployment environment: Mesos, VMware, Amazon Web Services (AWS), OpenStack, or another supported environment. Within the Infrastructure settings, the default cloud may be modified and additional clouds may be added.

Since each cloud is its own environment, networking and Service Engine (SE) settings are maintained separately within each cloud.

Dashboard

The default landing page for the Infrastructure section shows the dashboard for SEs. The SE dashboard display is similar to the one for virtual services (Applications > Dashboard), but shows only SEs.



All SEs across all clouds are shown. For each SE, the color indicates its health, with a numeric health score also shown next to the heart icon below the name of each SE. Hovering the mouse over the SE icon shows the SE health score breakdown. Clicking the SE name will jump to that SE’s page.

Clouds

Clouds are containers for the environment that Vantage is installed or operating within. During initial setup of Vantage, a default cloud, named “Default-Cloud”, is created. This is where the first

Controller is deployed, into Default-Cloud. Additional clouds may be added, containing SEs and virtual services.

If deploying redundant Controllers in a 3-node cluster, all 3 Controllers must be deployed into the same cloud.

A Controller is not limited to a single cloud, and also can belong to Controller clusters in other clouds.

The cloud table presents a list of the configured clouds.

- **Name:** The name of the cloud. The initial cloud is always named Default-Cloud.
- **Type:** Vantage may be installed in many types of environments, such as vCenter, OpenStack, or bare metal servers (no orchestrator). Vantage may be deployed into multiple environments, but only once per type. For example, a single Vantage deployment may connect to only one vCenter.
- **Status:** The colored status icon indicates the readiness of the cloud. Hovering the mouse over the icon provides more information about the status, such as ready for use or incomplete configuration.
- **Additional Icons:** The far right column of the table has a number of additional icons. The exact icons shown will depend on the clouds configured and their status.
    - **Edit:** Open the edit modal for the cloud.
    - **Convert:** Convert the cloud from read access mode or write access mode to no access mode. When in no access mode, Avi Controllers do not have access to the cloud's orchestrator, such as vCenter. See the installation documentation for the orchestrator to see the full implications of no access mode.
    - **Expand:** Click the plus icon or anywhere within the table row to expand the row and show more information about the cloud. For instance, in AWS the Region, Availability Zone, and Networks are shown.
    - **Download SE Image:** When Vantage is deployed in read access mode or no access mode, SEs must be installed manually. Use this button to pull the SE image for the appropriate image type (ova or qcow2). The SE image will have the Controller's IP or cluster IP address embedded within it, so an SE image may only be used for the Vantage deployment that created it.
    - **Generate Token:** Authentication tokens are used for securing communication between Controllers and SEs. If Vantage is deployed in read access mode or no access mode, the

SE authentication tokens must be copied manually by the Vantage user from the Controller web interface to the cloud orchestrator. For example, in a VMware deployment, the OVF template deployment dialog requires the Controller's authentication token as one of the input values.

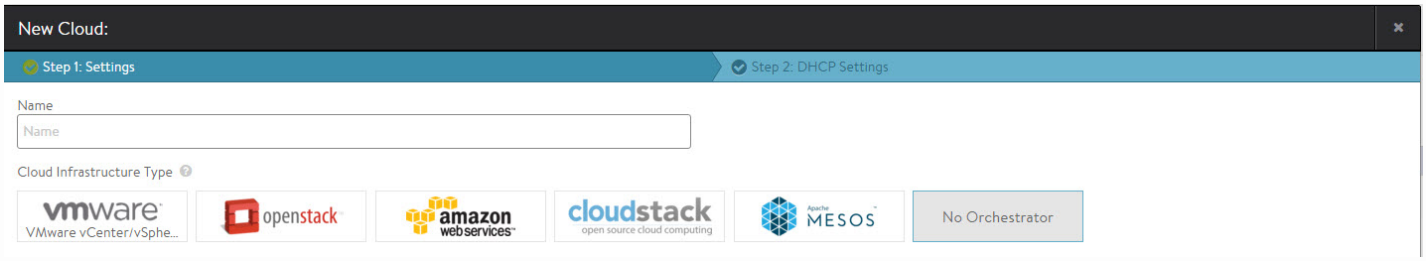
If needed for your read access mode or no access mode deployment, click this icon to display the Controller's authentication token, then copy-and-paste the token into the appropriate field in the cloud orchestrator interface. (See the Vantage installation guide for your infrastructure type for details.)

- **Install LBaaS Plugin:** For OpenStack clouds, this icon opens the LBaaS plugin dialog. Input the information regarding the Neutron server to install the plugin. Vantage will automatically push the LBaaS package to Neutron. If other considerations prevent using this method to install the plugin, such as requiring an SSH key instead of a Neutron password, then manually install by downloading the plugin from avinetworks.com/portal.

Cloud Creation

An initial cloud is created by default when Vantage is first deployed, and is documented in the Installation Guide for the appropriate cloud environment. To add an additional cloud, click the green New button to add the new cloud to the Vantage deployment.

Select the desired cloud. For deploying Avi SEs on bare metal servers, select No Orchestrator. This step is for the virtualization infrastructure and orchestrator. Supported network SDN technologies such as Cisco APIC or Nuage may be configured as additional properties of the cloud for which they are supported.



Vantage supports hybrid deployments. A single deployment can exist in multiple environments (for example, VMware and OpenStack) at the same time. Vantage does not support multiple clouds of the same type. For instance, Vantage cannot be deployed in write access mode into multiple VMware vCenter clouds.

Cloud Management

Clicking on the name of a cloud allows configuration of infrastructure objects within the environment. Each of these objects are specific to this cloud. For instance, a default static route configured in cloud 1 is only applicable to SEs in that cloud, and will not affect SEs in another

cloud.

Service Engines

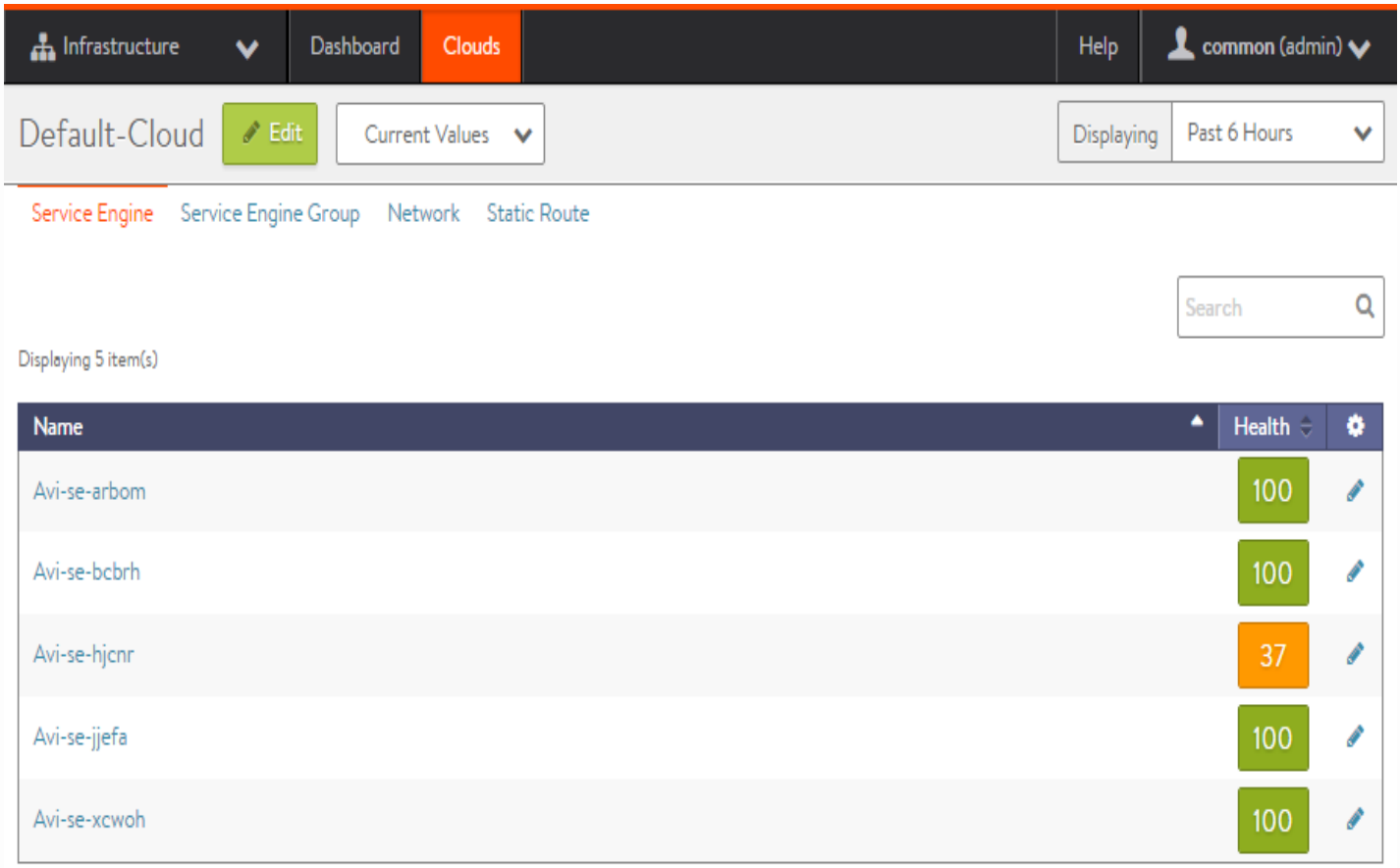
Avi Service Engines (SEs) handle all of the data plane operations within Vantage. SEs host the virtual services and require either direct or routable access to all client and server networks a virtual service touches.

A typical Vantage deployment may have many SEs for various purposes, such as redundancy, scalability, and accommodating large numbers of applications being served. SEs are always grouped within the context of a SE group, which provides settings for high availability, scalability, and potentially resource isolation for tenants.

- Service Engines Page ›
- Service Engine Quick Info Popup ›
- Create a Service Engine ›
- Delete a Service Engine ›

Service Engines Page

The Service Engines page lists the SEs that are currently configured in Vantage.



To display the SE list for a cloud, select Infrastructure > Clouds, click on a cloud name, and click Service Engines.

This page includes the following functions:

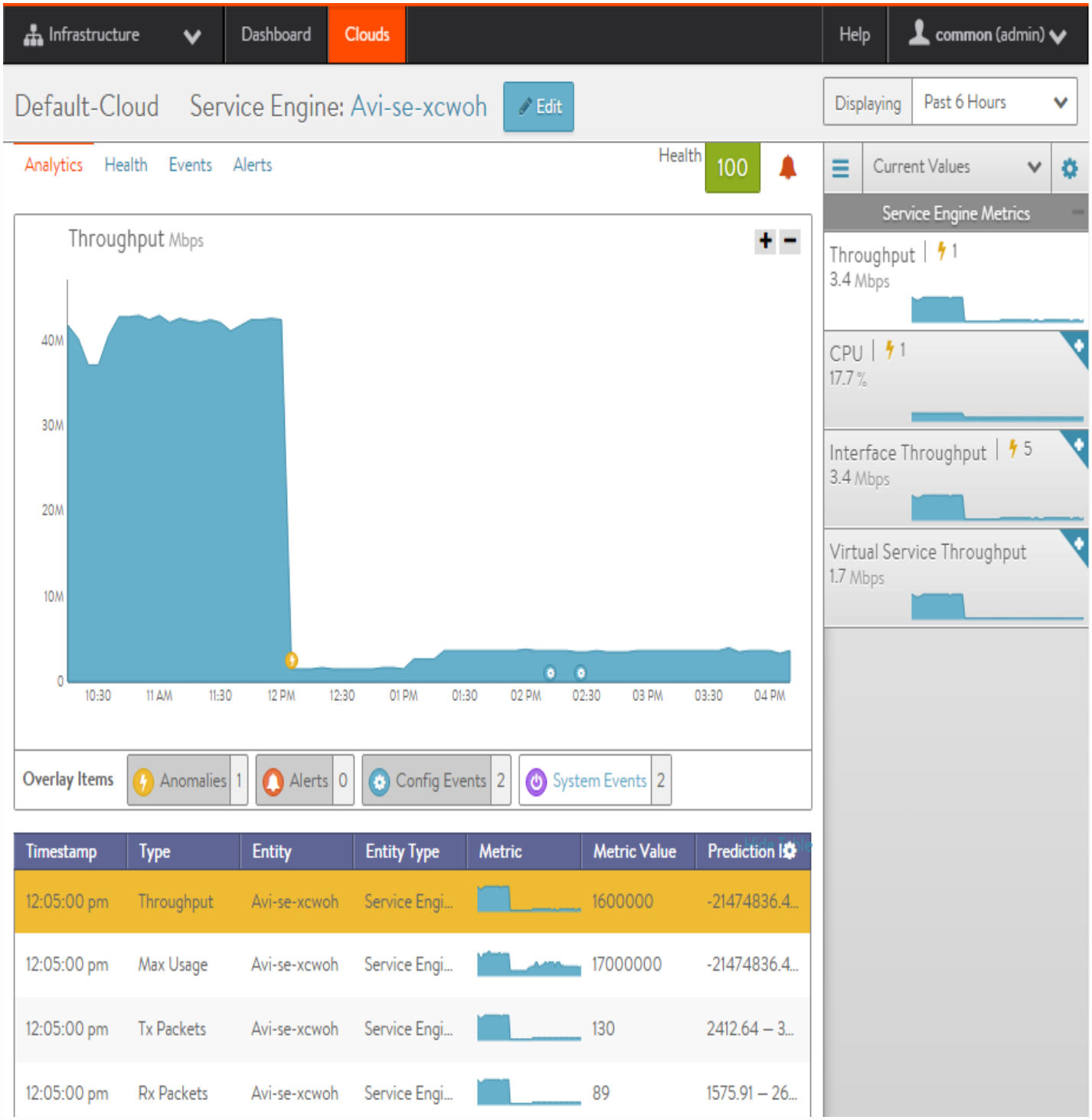
- **Search:** Search through the list of object names.
- **Edit:** Opens the Edit Service Engine popup.

This page contains the following information for each SE in the selected cloud:

- **Name:** Lists the name of each SE. Clicking the name of an SE opens the Analytics tab of the Service Engine Details page.
- **Health:** Provides both a numeric health score from 1-100 and a color-coded status to provide quick information about the health of the SE.
  - Hovering the cursor over the score opens the Health Score popup for the SE.
  - The View Health link at the bottom of the popup opens the Health tab of the Service Engine Details page.
  - Clicking within the Health Score opens the Analytics tab of the Service Engine Details page.

Service Engines Details Page

The Service Engine Details page shows information about the currently selected SE.

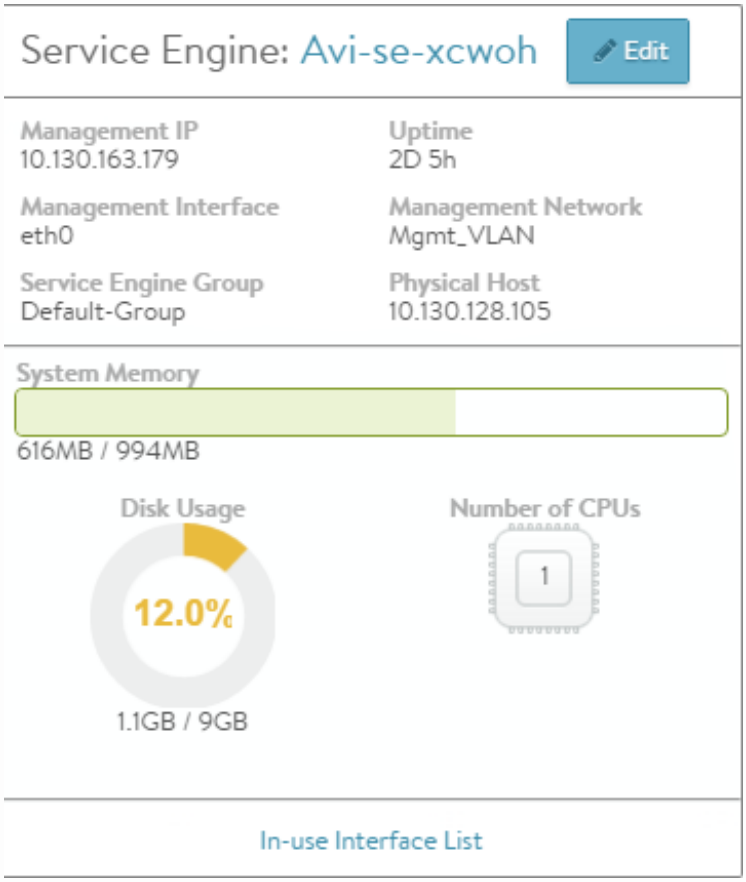


This page contains the following popup and tabs:

- Quick Info Popup ›
- Analytics Tab ›
- Health Tab ›
- Events Tab ›
- Alerts Tab ›

Service Engine Quick Info

Hovering over or clicking the name of the SE in the top left corner of the Service Engine Details page opens the Service Engine Info popup for that SE.



This popup provides the following information for the SE:

- **Management IP:** IP address the SE uses to communicate with the Controller.
- **Uptime:** The amount of time in days and hours that the SE has been either active or down.
- **Management Interface:** Network interface being used to allow the SE to communicate with the Controller. This address is reserved for management, and is not used for data plane or load balanced traffic. If management and data plane traffic will share the same network, they will still use two separate network interfaces and IP addresses.
- **Management Network:** Network used by the SE to communicate with the Controller. This may be the same network as one of the data networks used for load balancing. Best practice is to utilize a separate, dedicated network for control plane communications.
- **Service Engine group:** SE group that this SE belongs to. If you did not create an SE group, or the virtual service was not assigned to a unique SE group, then a new SE will default to the Default SE group.

- **Physical Host:** IP address of the physical server hosting the virtual machine on which the SE is running.
- **System Memory:** Amount of used versus available memory. Memory utilization should not exceed 90% for an extended period of time.
- **Disk Usage:** Percentage of allocated storage space being used by the SE. By default, an SE will be allocated 10 GB of storage. As the storage becomes full, logs may be purged prior to indexing. Adding more storage to a SE allows a greater volume of logs to be stored.
- **Number of CPUs:** Number of virtual CPU cores allocated to the SE. An idle SE will still consume some CPU as it is running normal housekeeping processes. An SE should not exceed 90% for an extended period of time as it may introduce latency in client transactions.

Service Engine Analytics

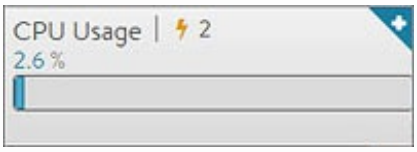
The Analytics tab presents information about various performance metrics over the time period selected.

Service Engine Analytics: Metrics

The following metrics are available for SEs:

- **Throughput:** Total bandwidth flowing through the SE for all virtual services being hosted by that SE. This includes the bandwidth flowing in and out of the SE between the client and the virtual service, and the traffic between the SE and the servers. Thus, an SE may report approximately double the throughput of its virtual services.

- **CPU Usage:** Displays the utilization of the CPUs allocated to the SE. The total number of CPUs appears in the Service Engine Quick Info Popup. Under normal conditions, CPU usage should not regularly exceed 90%, as this may cause latency in the virtual services and disrupt the client experience. The CPU Usage metric tile shows a horizontal bar indicating current usage, with a red line at the right to indicate how close the SE is to pushing the limits of its available CPU capacity.



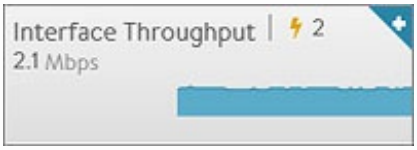
  You may indirectly control or improve CPU usage by taking actions, such as:

  - **Configuration:** Changing the configuration of virtual services, such as changing SSL or compression settings, will impact the CPU usage.
  - **CPU Allocation:** Allocating more vCPUs per SE. The default setting is two vCPUs per SE. Increasing this number is particularly useful for tasks such as SSL termination or

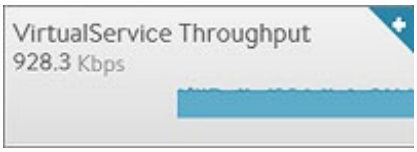
compression which heavily consume CPU resources. The setting for the number of vCPUs assigned to an SE is in the SE group.

- **Scale Out:** Reduce the CPU load by scaling this SE's virtual services across additional SEs, which will increase the total capacity and reduce the load on this SE. The high availability setting of the SE group dictates when a virtual service should be scaled out across additional SEs or simply migrated away from a busy SE.
- **CPU Reservation:** By default, CPUs resource is not reserved in a VMware deployment. Within vCenter, you may enable reservation for the SE's virtual machine, which guarantees that other virtual machines sharing the same physical host server will not be able to borrow or compete for CPU resources. This setting may be changed in the SE group properties. Changes will take effect for new SEs only. To make this change for existing SEs, it must be manually changed within vCenter. Refer to your VMware documentation.

- **Interface Throughput:** Shows the combined throughput for all network interfaces utilized by this SE. Throughput is measured as both client and server side of any virtual services, plus the management traffic between the SE and the Controllers.



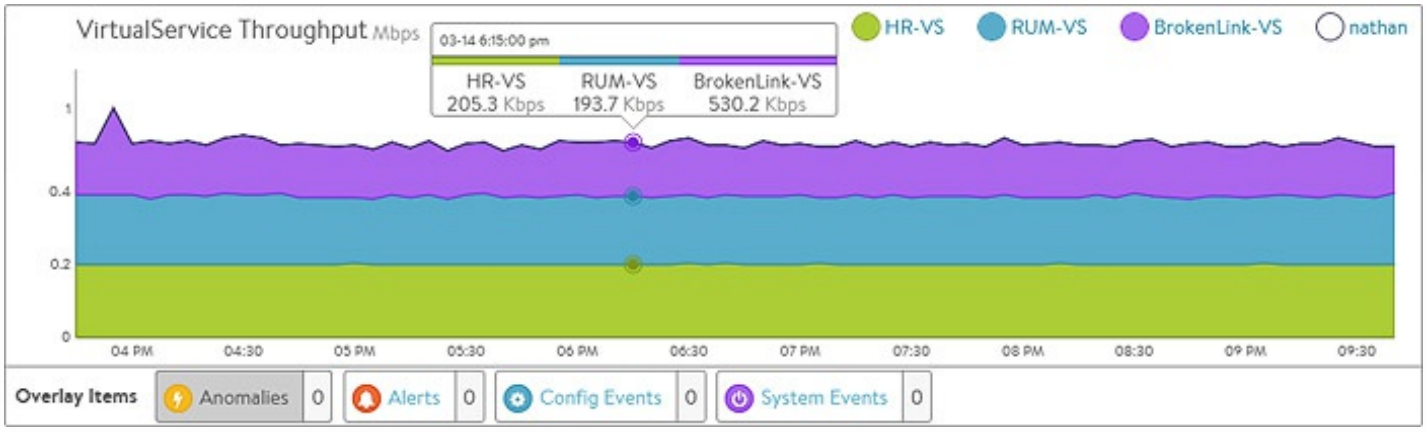
- **Virtual Service Throughput:** Shows the combined throughput for all network interfaces utilized by this SE. Throughput is measured as both client and server side of any virtual services, plus the management traffic between the SE and the Controllers.



Service Engine Analytics: Chart Pane

The main chart pane in the middle of the Analytics tab displays a detailed historical chart of the selected Metric tile for the current virtual service, pool, or SE.

- Hovering the mouse over any point in the chart will display the results for that selected time in a popup window.
- Clicking within the chart will freeze the popup at that point in time. This may be useful when the chart is scrolling as the display updates over time.
- Clicking again will unfreeze the highlighted point in time.

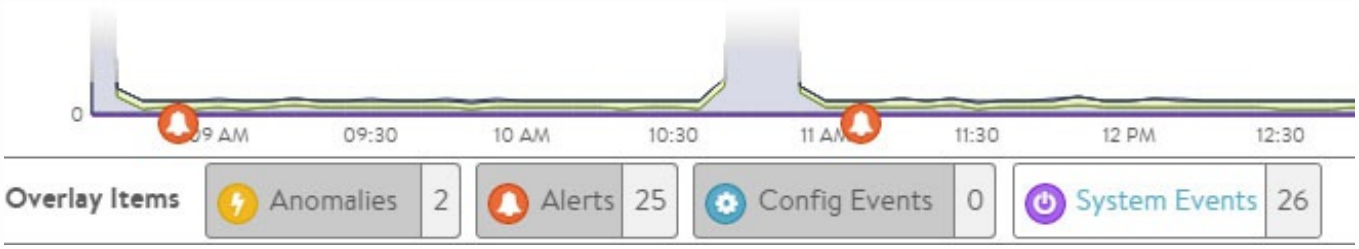


Many charts contain radio buttons in the top right that allow you to customize which data should be excluded from the chart. For example, if the End to End Timing chart is heavily skewed by one very large metric, then deselecting that metric by clearing the appropriate radio button will re-factor the chart based on the remaining metrics shown. This may change the value of the vertical Y-axis.

Some charts also contain overlay items, which will appear as color-coded icons along the bottom of the chart.

Service Engine Analytics: Overlays Pane

The overlays pane allows you to overlay icons signifying important events within the timeline of the chart pane. This feature helps you correlate anomalies, alerts, configuration changes, and system events with changes in traffic patterns.



Within the overlays pane:

- Each overlay type displays the number of entries for the selected time period.
- Clicking an overlay button toggles that overlay's icons in the chart pane. The button lists the number of instances (if any) of that event type within the selected time period.
- Selecting an overlay button displays the icon for the selected event type along the bottom of the chart pane. Multiple overlay icon types may overlap. Clicking the overlay type's icon in the chart pane will bring up additional data below the Overlay Items bar. The following overlay types are available:
    - **Anomalies:** Display anomalous traffic events, such as a spike in server response time,

along with corresponding metrics collected during that time period.

- **Alerts:** Display alerts, which are filtered system-level events that have been deemed important enough to notify an administrator.
- **Config Events:** Display configuration events, which track configuration changes made to Vantage by either an administrator or an automated process.
- **System Events:** Display system events, which are raw data points or metrics of interest. System events can be noisy, and are best used by alerts which filter and classify these raw events by severity.

SE Analytics: Anomalies Overlay

The Anomalies overlay displays periods during which traffic behavior was considered abnormal based on recent historical moving averages. Changing the time interval will provide greater granularity and potentially show more anomalies. Clicking the Anomalies Overlay button displays yellow anomaly icons in the chart pane, which can scroll down to view additional information related to that anomaly. During times of anomalous traffic, Vantage records any metrics that have deviated from the norm, which may provide hints as to the root cause of the anomaly.

:etoN
gniwoH

:etoN

These metrics are defined as follows:

Anomalies
(2014-10-20, 10:20:00 am – 2014-10-20, 11:10:00 am)
Hide Table

Timestamp	Type	Entity	Entity Type	Timeseries	Deviation
10:45:00 am	Data Transfer	Anomaly-VS	Virtual Service		Higher (**12.05**)
Other Relevant Anomalies					
10:40:00 am	CPU Usage	WebApp2	Virtual Machine		Lower (**-6.11**)
10:40:00 am	CPU Usage	WebApp1	Virtual Machine		Higher (**6.04**)

- **Timestamp:** Date and time when the anomaly was detected. This may either span the full duration of the anomaly, or merely be near the same time window.
- **Type:** The specific metric deviating from the norm during the anomaly period. To be included, the metric deviation must be greater than 4 sigma. Numerous types of metrics, such as CPU

utilization, bandwidth, or disk I/O may trigger anomalous events.

- **Entity:** Name of the specific object that is reporting this metric.
- **Entity Type:** Type of entity that caused the anomaly. This may be one of the following:
    - Virtual Machine (server); these metrics require Vantage to be configured for either read or write access to the virtualization orchestrator such as vCenter or OpenStack. In the example above, CPU utilization of the two servers was learned by querying vCenter.
    - Virtual service
    - SE
- **Time Series:** Thumbnail historical graph for the selected metric, including the most current value for the metric which will be data on the far right. Moving the mouse over the chart pane will show the value of the metric for the selected time. Use this to compare the normal, current, and anomaly time periods.
- **Deviation:** Change or deviation from the moving average, either higher or lower. The time window for the moving average depends on the time series selected for the Analytics tab.

SE Analytics: Alerts Overlay

The Alerts overlay displays the results of any events that meet the filtering criteria defined in the Alerts tab. Alerts notify administrators about important information or changes to a site that may require immediate attention.

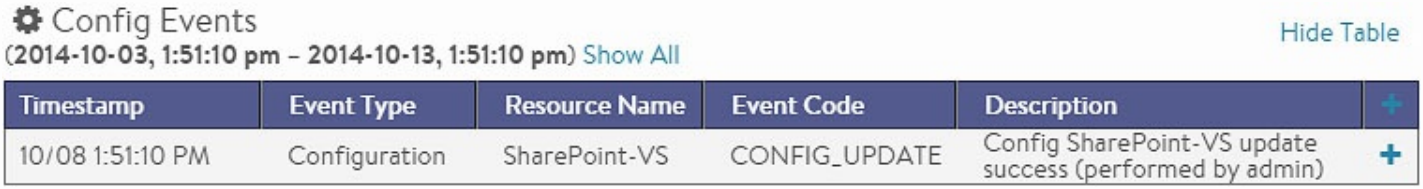
Alerts may be transitory, meaning they may expire after a defined period of time. For instance, Vantage may generate an alert if a server is down and then allow that alert to expire after a specified time period once the server comes back online. The original event remains available for later troubleshooting purposes.

Clicking the Alerts icon in the Overlay Items bar displays any red Alerts icons in the chart Pane. Selecting one of these chart alerts will bring up additional information below the Overlay Items bar, which will show the following information:

- **Timestamp:** Date and time when the Alert occurred.
- **Resource Name:** Name of the object that is reporting the Alert.
- **Level:** Severity of the Alert. You can use the priority level to determine whether additional notifications should occur, such as sending an email to administrators or sending a log to Syslog servers. The level may be one of the following:
  - **High:** Red
  - **Medium:** Yellow
  - **Low:** Blue
- **Summary:** Brief description of the event.
- **Actions:**
  - **Dismiss** the Alert with the red X to remove it from both the list shown and the Alert icon the chart pane. Dismissing an Alert here is the same as dismissing it via the bell icon at the top of the screen next to the Health Score, or dismissing it via the Alerts tab.
  - **Edit** the Alert filter to make Vantage more or less sensitive to generating new alerts.
- **Expand/Contract:** Clicking the plus (+) or minus sign (-) for an Alert opens and closes a sub-table showing more detail about the Alert. This will typically show the original events that triggered the alert.

SE Analytics: Config Events Overlay

The Config Events overlay displays configuration events, such as changing the Vantage configuration by adding, deleting, or modifying a pool, virtual service, or SE, or an object related to the object being inspected. If traffic dropped off at precisely 10:00 a.m., and at that time an administrator made a change to the virtual services security settings, there's a good chance the cause of the change in traffic was due to the (mis)configuration.



Clicking the Config Events icon in the Overlay Items bar displays any blue Config Event icons in the chart pane. Selecting one of these chart alerts will bring up additional information below the Overlay Items bar, which will show the following information:

- **Timestamp:** Date and time when the configuration change occurred.

- **Event Type:** Always be scoped to *Configuration* event types.
- **Resource Name:** Name of the object that has been modified.
- **Event Code:** There are three event codes:
  - CONFIG_CREATE
  - CONFIG_UPDATE
  - CONFIG_DELETE
- **Description:** Brief description of the event.
- **Expand/Contract:** Clicking the plus (+) or minus sign (-) for a configuration event either expands or contracts a sub-table showing more detail about the event. When expanded, this shows a difference comparison of the previous configuration versus the new configuration, as follows:
  - Additions to the configuration, such as adding a health monitor, will be highlighted in green in the new configuration.
  - Removing a setting will be highlighted in red in the previous configuration.
  - Changing an existing setting will be highlighted in yellow in both the previous and new configurations.

SE Analytics: System Events Overlay

This overlay displays System Events relevant to the current object, such as a server changing status from up to down or the health score of a virtual service changing from 50 to 100.

Clicking the System Events icon in the Overlay Items bar displays any purple System Event icons in the chart pane. Select a system event icon in the chart pane to bring up more information below the Overlay Items bar.

- **Timestamp:** Date and time when the system even occurred.
- **Event Type:** This will always be System.
- **Resource Name:** Name of the object that triggered the event.
- **Event Code:** High-level definition of the event, such as VS_Health_Change or VS_Up.
- **Description:** Brief description of the system event.

- **Expand/Contract:** Clicking the plus (+) or minus sign (-) for a system event expands or contracts that system event to show more information.

Service Engine Health

The health score of an on object is comprised from the following scores:

- **Performance:** Performance score (1-100) for the selected item. A score of 100 is ideal, meaning clients are not receiving errors and connections or requests are quickly returned.
- **Resource Penalty:** Any penalty assessed because of resource availability issues is assigned a score, which is then subtracted from the Performance score. A score of 0 is ideal, meaning there are no obvious resource constraints on Vantage or virtualization orchestrator connected servers.
- **Anomaly Penalty:** Any penalty assessed because of anomalous events is assigned a score, which is then subtracted from the Performance score. An ideal score is 0, which means Vantage has not seen recent anomalous traffic patterns that may imply future risk to the site.
- **Health Score:** The final health score for the selected item equals the Performance Score minus the Resource and Anomaly Penalty scores.

The sidebar tiles show the scores of each of the three subcomponents of the health score, plus the total score. To determine why an object such as a virtual service has a low health score, select one of the first three tiles that is showing a subpar score.

This will bring up additional sub-components for the top level metric, such as pools, connection Apdex, Response Apdex, or others. Again, select the tile that is showing the worst score. Some tiles may have additional information shown in the main chart section that requires scrolling down to view. Clicking on a tile for another object such as a pool or SE will jump to the Insights page for that object.

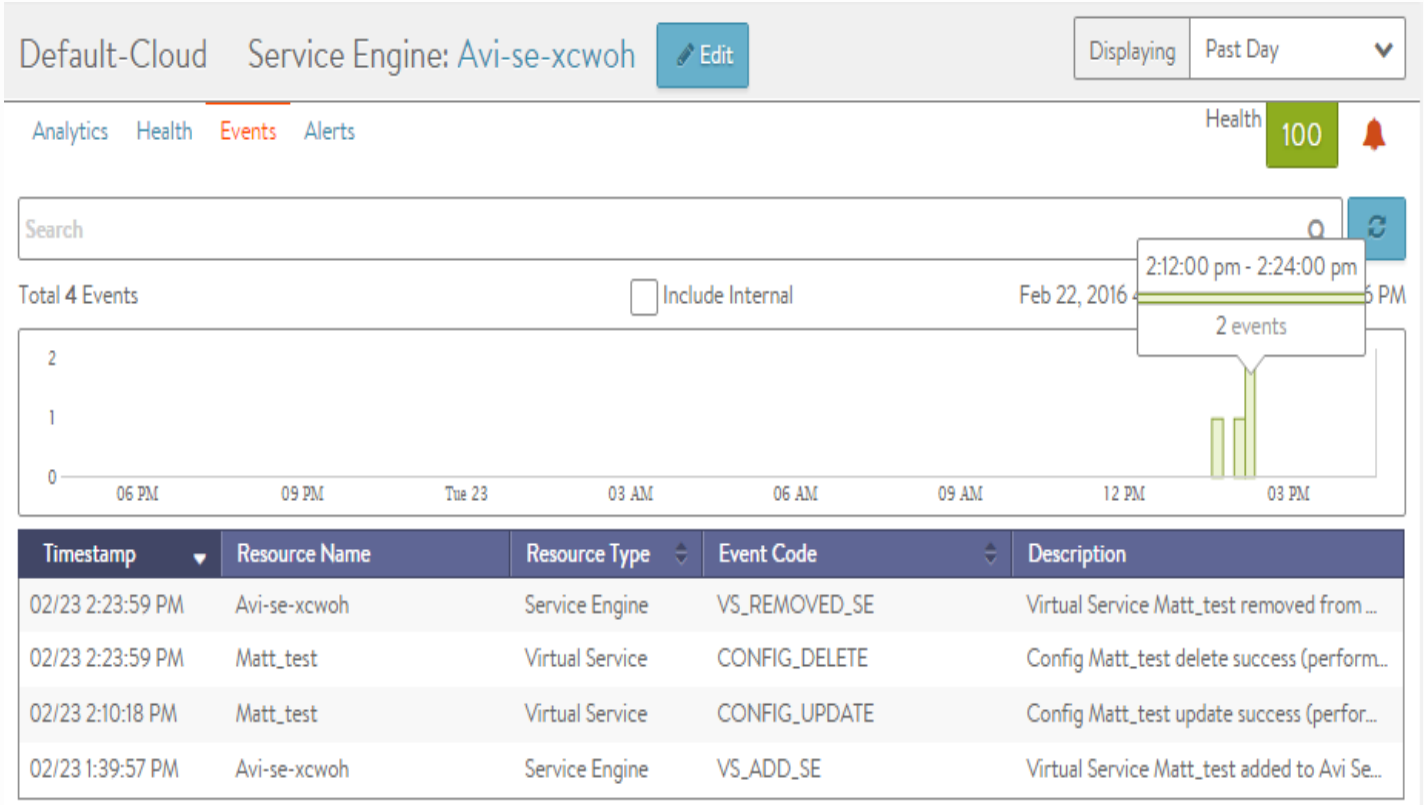
The chart pane of the tab shows a detailed graph of the selected score:

- Clicking any of the summary Metrics tiles on the sidebar displays the detailed version of that graph in the chart pane of the tab. Additional details may display at the bottom of the tab that show various factors contributing to that score.
- Hovering your mouse cursor over any of the charts displays the health score for the selected date and time on all graphs.

Service Engine Events

The Events tab presents system-generated events over the time period selected for the SE.

System events are applicable to the context in which you are viewing them. For example, when viewing events for a SE, only events that are relevant to that object are displayed.



The top of this tab displays the following items:

- **Search:** The Search field allows you to filter the events using whole words contained within the individual events.
- **Refresh:** Clicking Refresh updates the events displayed for the currently-selected time.
- **Number:** The total number of events being displayed. The date/time range of those events appear beneath the Search field on the left.
- **Clear Selected:** If filters have been added to the Search field, clicking the Clear Selected (**X**) icon on the right side of the search bar will remove those filters. Each active search filter will also contain an **X** that you can click to remove the specific filter.
- **Histogram:** The Histogram shows the number of events over the period of time selected. The X-axis is time, while the Y-axis is the number of events during that bar's period of time.
    - Hovering the cursor over a Histogram bar displays the number of entries represented by that bar, or period of time.
    - Click and drag inside the histogram to refine the date/time period which further filters the events shown. When drilling in on the time in the Histogram, a Zoom to selected link appears above the Histogram. This expands the drilled in time to expand to the width of

the Histogram, and also changes the Displaying pull-down menu to Custom. To return to the previously selected time period, use the Display pull-down menu.

The table at the bottom of the **Events** tab displays the events that matched the current time window and any potential filters. The following information appears for each event:

- **Timestamp:** Date and time the event occurred. Highlighting a section of the histogram allows further filtering of events within a smaller time window.
- **Event Type:** This may be one of the following:
  - **System:** System events are generated by Vantage to indicate a potential issue or create an informational record, such as VS_Down.
  - **Configuration:** Configuration events track changes to the Vantage configuration. These changes may be made by an administrator (through the CLI, API, or GUI), or by automated policies.
- **Resource Name:** Name of the object related to the event, such as the pool, virtual service, SE, or Controller.
- **Event Code:** A short event definition, such as Config_Action or Server_Down.
- **Description:** A complete event definition. For configuration events, the description will also show the username that made the change.
- **Expand/Contract:** Clicking the plus (+) or minus sign (-) for an event log either expands or contracts that event log. Clicking the + and – icons in the table header expands and collapses all entries in this tab.

For configuration events, expanding the event displays a difference comparison between the previous and new configurations.

- New fields will appear highlighted in green in the new configuration
- Removed fields will appear highlighted in red.
- Changed fields will show highlighted in yellow.

Service Engine Alerts

The Alerts tab displays specified events that have trigger an alert. Alert actions can be configured, and proactive notifications generated via Syslog or email in the Notifications tab of the Administration page. Alerts act as filters that provide notification for prioritized events or combinations of events through various mechanisms such as the Vantage web interface, email,

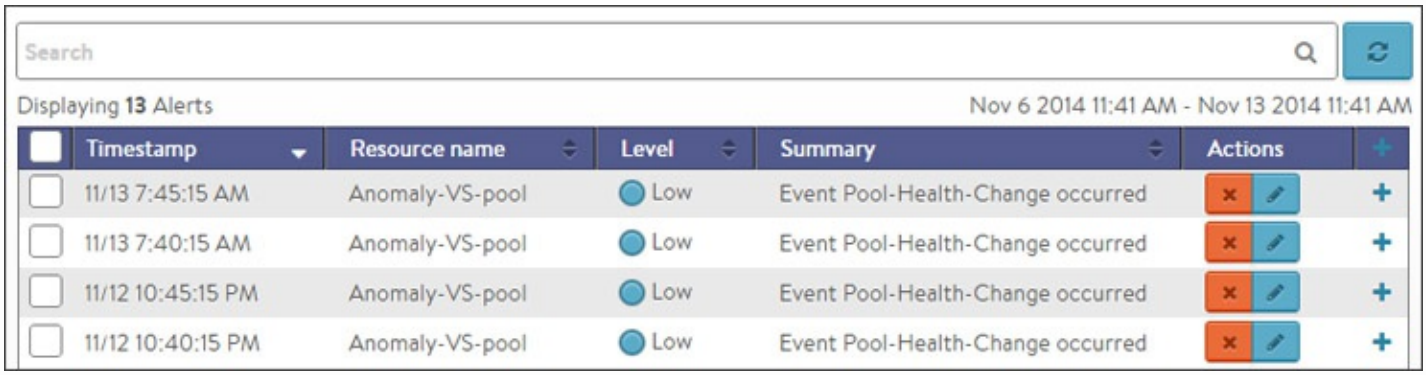
or Syslog. Vantage includes a number of default alerts based on events deemed to be universally important.

The top of this tab shows the following items:

- **Search:** The Search field allows you to filter the alerts using whole words contained within the individual alerts.
- **Refresh:** Clicking Refresh updates the alerts displayed for the currently-selected time.
- **Number:** The total number of alerts being displayed. The date/time range of those alerts appear beneath the Search field on the left.
- **Dismiss:** Select one or more alerts from the table below then click Dismiss to remove the alert from the list.

:etoN
f i ton
) desab

The table at the bottom of the Alerts tab displays the following alert details:



- **Timestamp:** Date and time when the alert was triggered. Changing the time interval using the Displaying pull-down menu may potentially show more alerts.
- **Resource Name:** Name of the object that is the subject of the alert, such as a Server or virtual service.
- **Level:** Severity level of the alert, which can be High, Medium, or Low. Specific notifications can be set up for the different levels of alerts via the Administration page's Alerts Overlay.
- **Summary:** Summarized description of the alert.
- **Action:** Click the appropriate button to act on the alert:
    - **Dismiss:** Clicking the red **X** dismisses the alert and removes it from the list of displayed

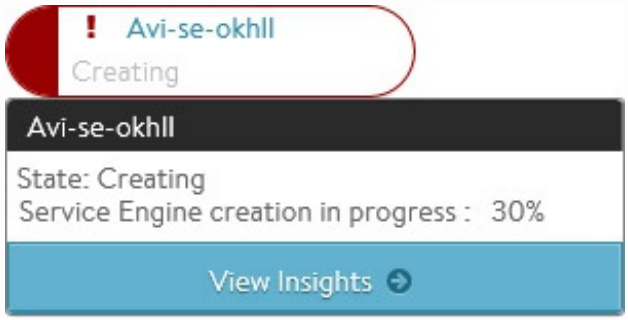
alerts.

- **Edit:** Clicking the blue pencil icon opens the Edit Alert Config popup for the alert configuration that triggered this alert. This can include a verbose and customized description of the alert or allow an administrator to alter settings such as the severity of the alert.

- **Expand/Contract:** Clicking the plus (**+**) or minus sign (**-**) for an event log either expands or contracts that event log to display more information. Clicking the **+** and **–** icon in the table header expands and collapses all entries in this tab

Service Engine Create: Write Access Mode Deployments

An Avi Controller that is deployed in write access mode has full write access to the virtualization platform and can automatically deploy new SEs and modify the network configuration of existing SEs. The Controller will place the virtual service on a SE within a cluster and host that has optimal reachability to the servers. In a new Vantage deployment, the first SE will not be created until the first virtual service is created, as this is required to know which server network will be used.



The health score of a newly created virtual service will appear as gray with an exclamation point while the SE is being created; hovering the mouse over the health score will show the status as Creating. During this time, the Controller copies the SE image file from itself to the host server, sets up virtual machine settings via the virtualization orchestrator, then sets the network adapters and IP addresses required to reach clients and servers. This process may take anywhere from a few seconds to a few minutes, depending on the time it takes to copy the SE image across the network to a physical host. If creation of the SE fails, the Controller will wait for five minutes and then attempt to recreate the SE on a new host.

In an established environment, a new virtual service may use an existing SE and thus be brought up immediately. Preferences for high availability, scalability, and number of virtual services per SE are defined within the SE group settings.

If all virtual services for a SE are deleted and the SE is no longer in use, the Controller will wait 120 minutes before automatically removing the unused SE. This setting may be configured via the SE group properties.

Service Engine Create: Read/No Access Mode Deployments

When Vantage is deployed in read access mode or no access mode, Vantage does not have write access to the virtualization infrastructure. In this case, an administrator must manually perform any operations that require write access to the virtualization environment (create and delete SEs and configure network settings).

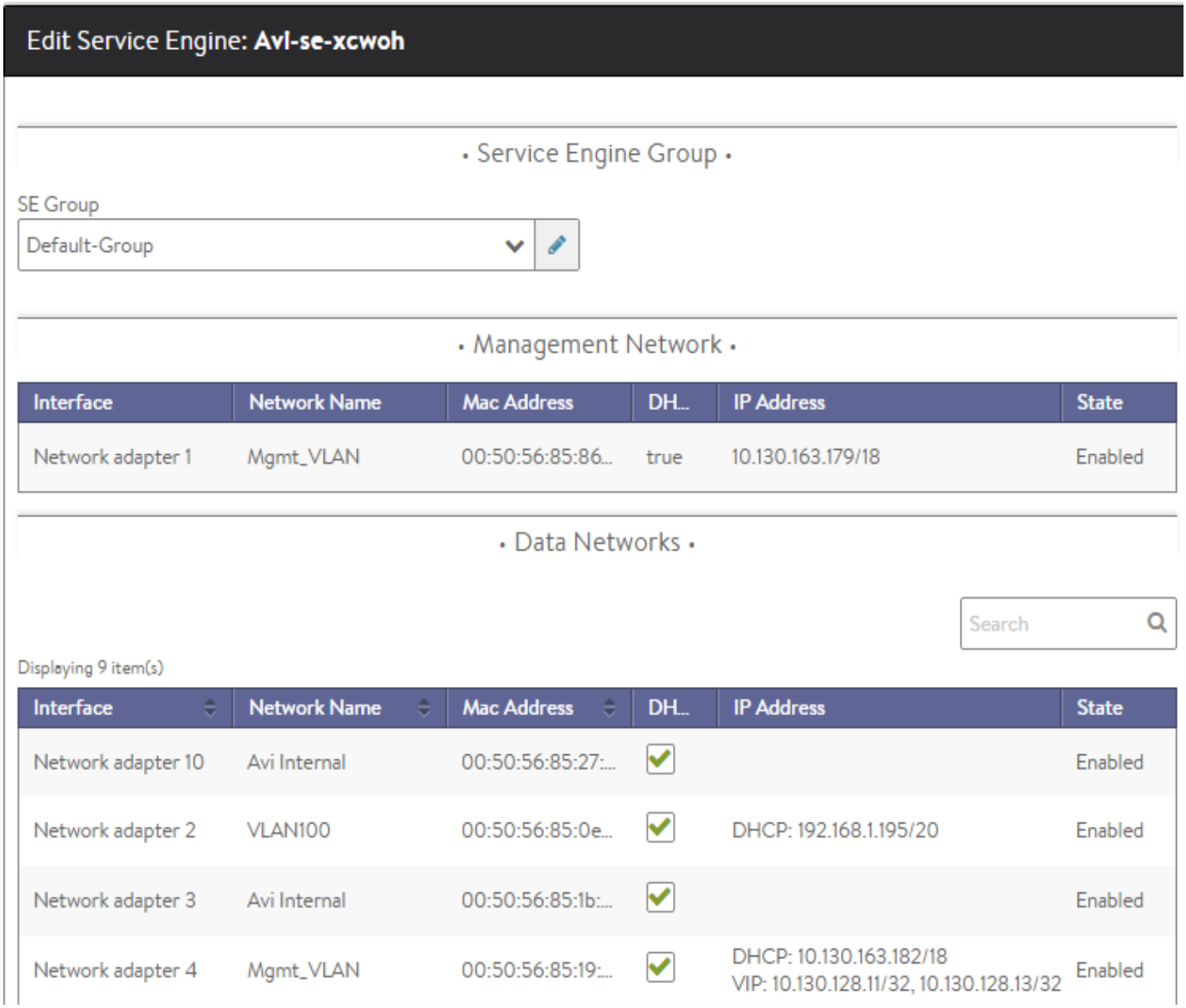
A new virtual service may be able to use an existing SE, though it may still require an administrator to change the network settings such as adding a new network interface into a port group required for access to servers.

Creating a new SE when the Avi Controller has Read or no access to the virtualization platform is almost identical to the process described in the Installation Guide for your selected virtualization platform, except that:

- If the data plane network interfaces (those processing load balanced traffic) need to be set to a static IP address, an administrator will need to manually match the network interface shown in the Avi Controller with the Network Adapter shown in the virtualization platform. The Controller cannot poll the Network Name because it does not have access to the virtualization platform.
- An admin will need to find the MAC Address of the virtual machine’s network adapter that clients wish to use, and then correspond that to the MAC Address shown in the Edit a Service Engine popup.

Edit a Service Engine

The Edit Service Engine popup allows an administrator to modify the network settings for the SE. To edit an SE, select Infrastructure > Service Engines and click on the SE name or on the edit icon.



Many networking properties can be configured on the Networks tab and in the Service Engine Edit popup. Editing the SE properties will only affect the specific SE being modified; you will need to manually modify any new SE created thereafter. If Vantage has No access to the hypervisor, the administrator will need to manually edit the network and IP settings for each SE. For deployments in write access mode, editing the values on the Network tab is needed to ensure that any new SE will inherit the desired settings.

- **Service Engine Group:** An SE may be manually migrated to a different SE group by selecting the new SE group from the dropdown menu. Moving a SE is not graceful. It will first terminate any existing connections.
- **DHCP:** DHCP may be enabled per network interface, not per IP network. This is the default

setting for all network interfaces. An SE attempting to use DHCP to acquire an IP address will retry every five minutes and will generate an error in the events log if it is unsuccessful.

Note: A single interface may have multiple networks configured. It is therefore possible to have both DHCP and static IP addresses configured for a single interface.

- **Default Gateway:** Enter a new IP address for the gateway in the Default Gateway field.

Delete a Service Engine

An SE may be deleted for many reasons, such as:

- Placement on a different physical host.
- Updating resource sizes (e.g. number of vCPUs)
- Reduced load no longer requires as many SEs.

If Vantage is deployed to have write access mode to the hypervisor orchestrator, Vantage will automatically delete unused SEs. If Vantage is deployed in read access mode or no access mode, SEs may be deleted from the Controller, but it will still require an administrator to manually delete the SE from the virtualization platform.

Note: To delete an SE from a Controller immediately rather than wait for the SE to time out based on the SE group settings, use the CLI or API.

Service Engine Group

An SE group is a collection of one or more SEs that may share properties, such as network access and failover. An SE cannot scale out across or fail over to an SE which is in a different SE group, even if both SEs share the same physical host or network properties. Different applications can thus receive guaranteed data plane isolation when deployed on different SE groups.

Virtual services created in a new Vantage deployment will be assigned to the Default-Group SE group. To deploy virtual services to a different SE group:

1. Create a new SE group.
2. Move or create the new virtual service in the new group using the Advanced tab of the Edit Virtual Service page.

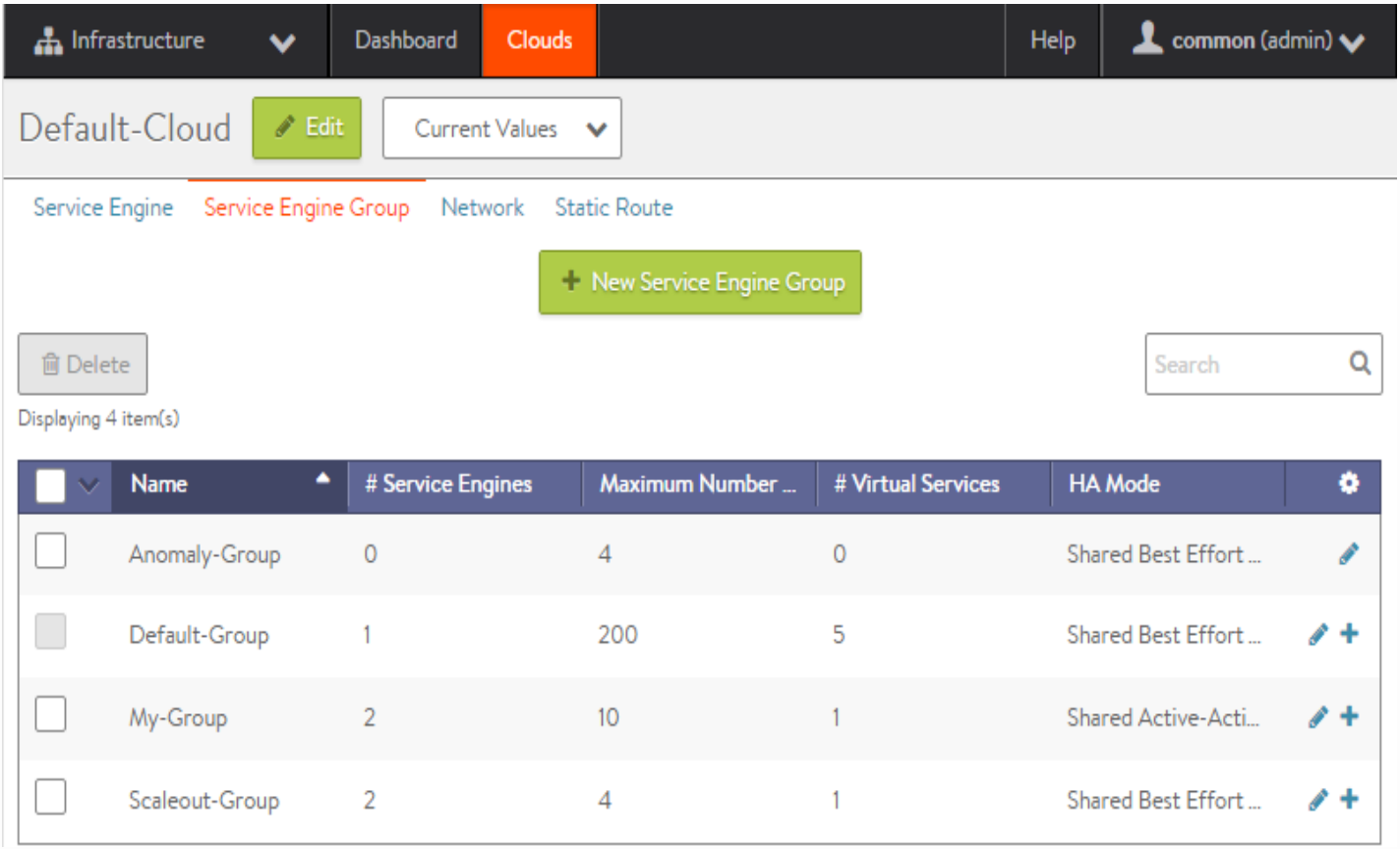
When creating a new SE group in write access mode, no new SEs will be created until a virtual service is deployed to the SE group. In read access mode or no access mode deployments, the new SEs must be manually created. They will attempt to connect back to the Controller after they have booted up, at which point they will be added to the Default SE group. SEs in read access

mode and no access mode deployments can be migrated to a new SE group, provided all virtual services deployed on the SE are disabled.

SEs in write access mode deployments cannot be migrated to new SE groups. Instead, the old SE is deleted and a new SE is created. This process is automatic if the virtual services are migrated.

Service Engine Groups Page

The Service Engine Groups page lists the configured SE groups.



The table on this page contains the following information for each SE group:

- **Name:** Lists the name of each SE group.
- **# Service Engines:** Shows the number of SEs assigned to the SE group. Clicking the row on the table will show an expanded view with the names of SEs.
- **Maximum Number of Service Engines:** Maximum number of SEs the group can contain.
- **# Virtual Services:** Shows the number of virtual services assigned to the SE group. Clicking the row on the table will show an expanded view with the names of virtual services.
- **HA Mode:** High availability mode configured for the group.

Note: Only unused SE groups may deleted. If the SE group is in use by a virtual service, a popup will warn that dependent virtual services must first be deleted or migrated to other SE groups via the Virtual Service > Edit > Advanced properties tab. A tenant must always have a minimum of one configured SE group. The default SE group may be modified, but not deleted.

Create a Service Engine Group

To create or edit an SE group:

1. Select Infrastructure > Clouds and click on the cloud name (for example, Default-Cloud).
2. Select Service Engine Group to open the Service Engine Groups page, which lists the SE groups currently configured in Vantage.
3. Click New Service Engine Group or click on an SE group anem in the table.

The create and edit popups for SE groups have identical properties. This popup includes the following tabs:

- High Availability Tab
- Advanced Tab

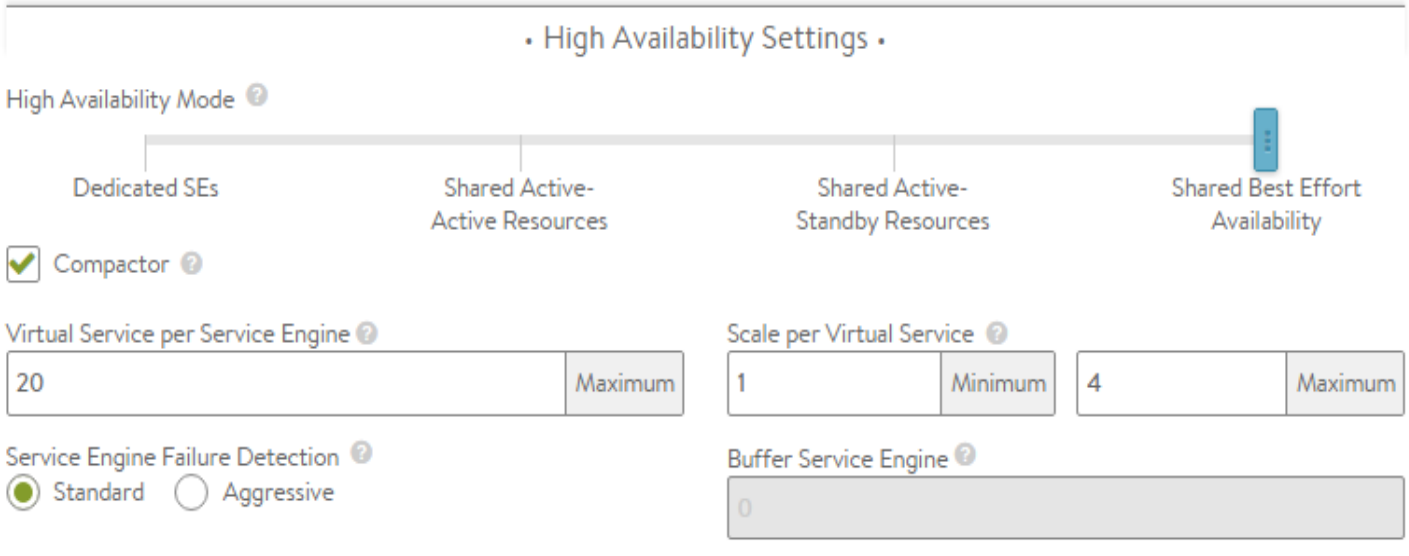
High Availability Tab

Click New in the Edit Service Engine Group popup to open the High Availability tab.

Edit the High Availability settings:

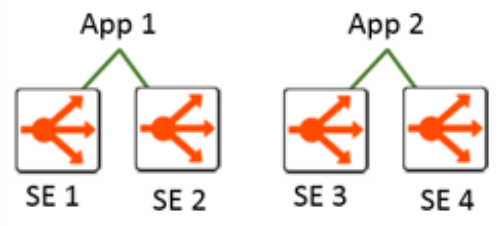
- Name: Enter a unique name for the SE group in the Name field.
- Optionally configure any setting within the High Availability tab.
- Either click the Advanced Tab, or Save to return to the Service Engine Groups page.

Virtual Service High Availability Settings



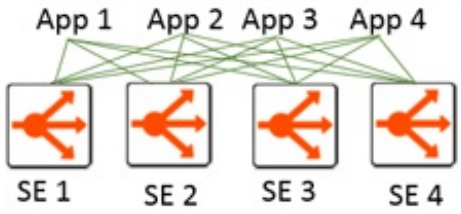
- **High Availability Mode:** Controls how load is scaled across SEs and how Vantage handles SE redundancy and failures. Switching between the four HA modes will change the settings and options exposed below the slider bar. These modes provide a range of choices for using the fewest virtual machine resources versus providing the best high availability.
- **Dedicated SEs:** Every virtual service IP address (which may be multiple virtual services on different ports, but sharing the same IP address) is allocated a redundant, dedicated pair of SEs and will be scaled across both (active/active). If an SE fails, the remaining SE will remain active. Maximum capacity will be halved while Vantage creates a new SE and replaces the failed SE.

App	Primary SE	Active SE	Standby SE
WebApp1	SE 1	SE 1, 2	n/a
WebApp2	SE 3	SE 3, 4	n/a



- **Shared Active-Active Resources:** Every new virtual service will be striped across at least two SEs. This allows both better scale, but also faster failover times. If one of the SEs fails, capacity is temporarily reduced while another SE is configured or a new SE is created to backfill the down SE. This elastic mesh ensures maximum availability for virtual services while also ensuring that all SEs are utilized.

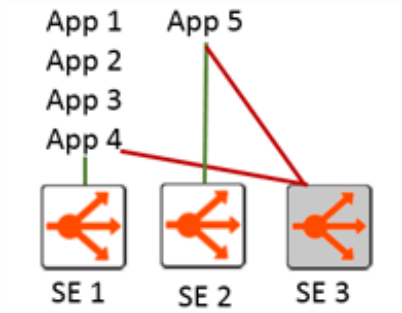
App	Primary SE	Active SE	Standby SE
WebApp1	SE 1	SEs 1, 2	n/a



WebApp2	SE 2	SEs 1, 2	n/a
WebApp3	SE 3	SEs 3, 4	n/a
WebApp4	SE 4	SEs 3, 4	n/a

- **Shared Active-Standby Resources:** With this default mode, new virtual services will initially be stacked on a single SE until it reaches either resource capacity or the Maximum Virtual Services per SE setting. One extra, standby SE will also be created. In this mode, the SE group will always have the equivalent of one SE worth of capacity available. The illustration shows the distribution of virtual services across a new SE group. However, as virtual services grow and shrink in capacity, they may be scaled across any of the SEs. No SE is a strict standby, rather they may all be active, and if the system does not have enough capacity to absorb the loss of an individual SE, then a new one will be proactively created. On a busy system that has been active for some time, the virtual services may begin to look more like the illustration from the Shared Active/Active setting. The biggest difference between these two modes is Active-Standby starts with the Compact mode, stacking virtual services on fewer SEs until the load dictates otherwise, and the system will try to maintain the buffer of one SE worth of unused, available capacity.

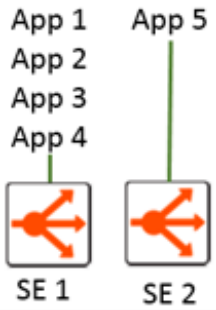
App	Primary SE	Active SE	Standby SE
WebApp1	SE 1	SE 1	SE 3
WebApp2	SE 1	SE 1	SE 3
WebApp3	SE 1	SE 1	SE 3
WebApp4	SE 1	SE 1	SE 3
WebApp5	SE 2	SE 2	SE 3



- **Shared Best Effort Availability:** This mode is similar to the Shared Active/Standby mode, except Vantage does not create a standby SE until an active SE fails. This will generally incur an outage of 30-90 seconds; however, creation time will depend on the speed of file copy and server bandwidth. This mode attempts to use the smallest possible number of virtual machines for SEs, but this does come at the cost of having the slowest recovery time for a SE failure.

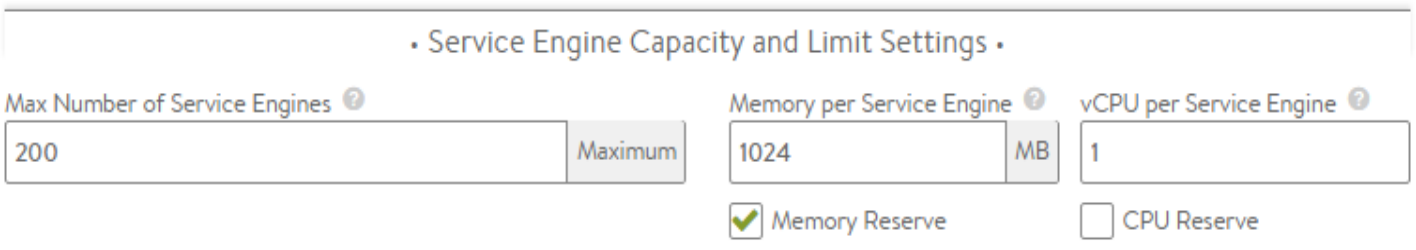
App	Primary SE	Active SE	Standby SE

WebApp1	SE 1	SE 1	n/a
WebApp2	SE 1	SE 1	n/a
WebApp3	SE 1	SE 1	n/a
WebApp4	SE 1	SE 1	n/a
WebApp5	SE 2	SE 2	n/a



- **Active Standby:** When enabled, each active SE will have a single corresponding standby SE. Failure of the active SE will cause the standby to take over as active. A new standby may be created if Vantage cannot recover the failed SE.
- **Compactor:** When enabled, new virtual services are placed on existing SEs with other virtual services. Disabling this option places each new virtual service in its own SE until the maximum number of SEs for the SE group is reached. At this point, a new virtual service will be placed on the SE with the least utilization. When this option set, Vantage will attempt to conservatively create new SEs.
- **Virtual Service per Service Engine:** Controls the maximum number of virtual services that may be deployed on a single SE. Another SE must be created or used if this maximum is reached. If Vantage reaches the maximum number of SEs, then no more virtual services can be deployed within the SE group.
- **Scale per Virtual Service - Minimum:** The virtual service may be scaled across multiple SEs, which both increases potential capacity and ensures recovery from any failure while minimizing impact. Setting the minimum above 1 ensures that every virtual service starts out scaled across multiple SEs regardless of capacity requirements.
- **Scale per Virtual Service - Maximum:** Sets the maximum number of SEs across which a virtual service may be scaled.
- **Buffer Service Engine:** This option only exists for Active/Standby modes when the Compactor setting is True. The SE group will create an additional, dedicated standby SE for the group that will be used to replace any failed SE.

Service Engine Capacity and Limit Settings



- **Number of Service Engines:** defines the maximum SEs that may be created within a SE group. This number, combined with the virtual services per SE setting, dictate the maximum number of virtual services that can be created within an SE group. If this limit is reached, it is possible new virtual services may not be able to be deployed and will show a gray, undeployed status. This setting can be useful for limiting Vantage from consuming too many virtual machines.

- **Memory per Service Engine:** Enter the amount of RAM, in MB, to allocate to all new SEs. Changes to this field will only affect new SEs. Allocating more memory to an SE will allow larger HTTP cache sizes, more concurrent TCP connections, better protection against certain DDoS attacks, and increased storage of un-indexed logs. This option is only applicable in write access mode deployments.

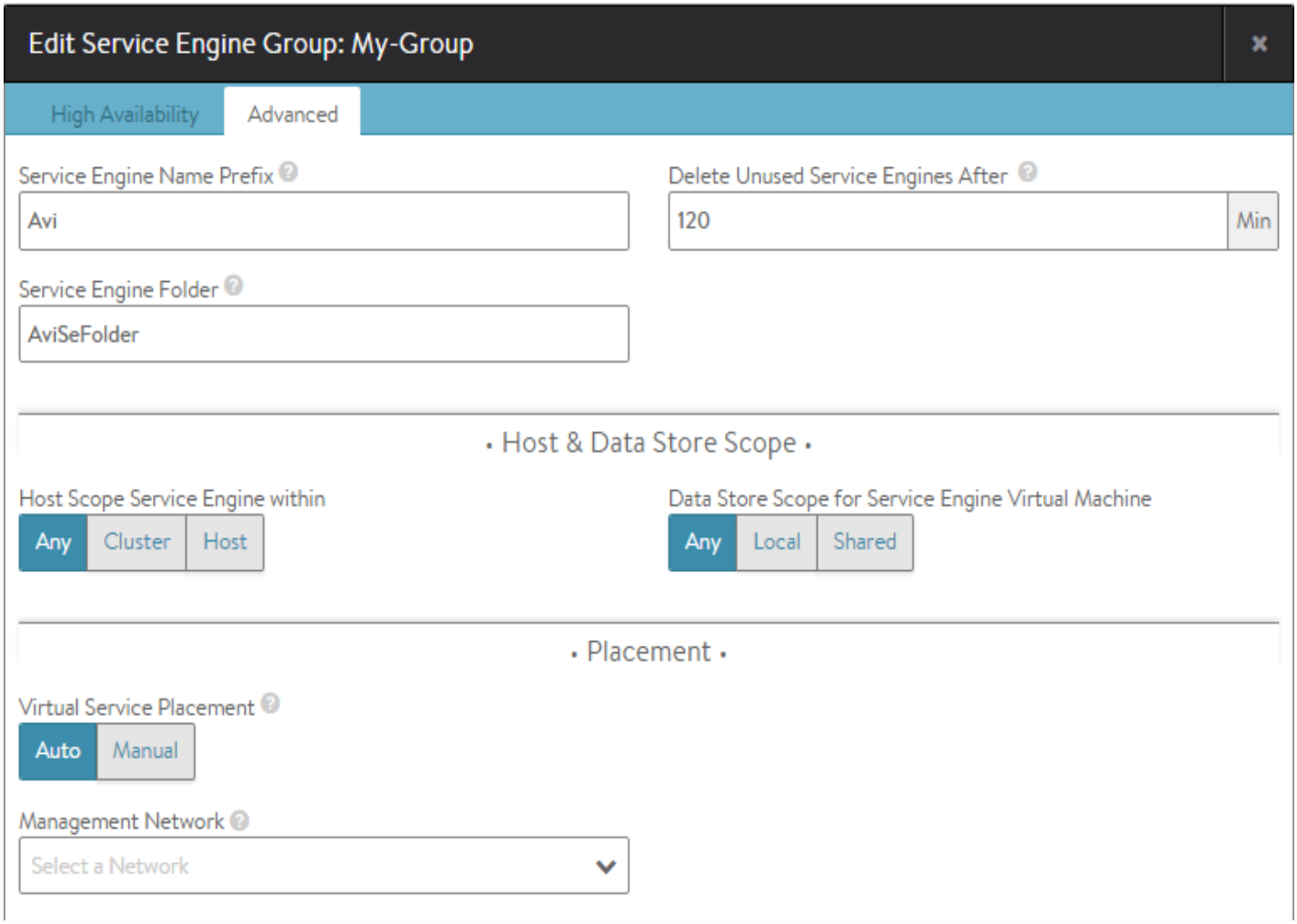
    - **Memory Reserve:** Reserving memory ensures an SE will not have contention issues with over-provisioned host hardware. Reserving memory makes that memory unavailable for use by another virtual machine, even when the virtual machine that reserved those resources is powered down. Avi recommends reserving memory, as memory contention may randomly overwrite part of the SE memory, destabilizing the system. This option is applicable only for deployments in write access mode. For deployments in read access mode deployments or no access mode, memory reservation for the SE VM must be configured on the virtualization orchestrator.
- **vCPU per Service Engine:** Enter the number of virtual CPU cores to allocate to new SEs. Changes to this setting do not affect existing SEs. This option is only applicable in write access mode. Adding CPU capacity will help with computationally expensive tasks, such as SSL processing or HTTP compression.

    - **CPU Reserve:** Reserving CPU capacity with a virtualization orchestrator ensures a SE will not have issues with over-provisioned host hardware. Reserving CPU cores makes those cores unavailable for use by another virtual machine, even when the virtual machine that reserved those resources is powered down. This option is only applicable in write access mode deployments.

Advanced Service Engine Group Settings

The Advanced tab in the Edit Service Engine Group popup allows configuration of optional functionality for SE groups. This tab appears only when Avi Vantage is deployed in write access mode deployments.

Note: This tab appears only when Vantage is deployed in write access mode.



- **Service Engine Name Prefix:** Enter the prefix to use when naming the SEs within the SE group. This name will be seen both within Vantage, and as the name of the virtual machine within the virtualization orchestrator.
- **Service Engine Folder:** SE Virtual Machines for this SE group will be grouped under this folder name within the virtualization orchestrator.
- **Delete Unused Service Engines After:** Enter the number of minutes to wait before the Controller deletes an unused SE. Traffic patterns can change quickly, and a virtual service may therefore need to scale across additional SEs with little notice. Setting this field to a high value ensures that Vantage keeps unused SEs around in case of a sudden spike in traffic. A shorter value means the Controller may need to recreate a new SE to handle a burst of traffic, which may take a couple of minutes. This option is only applicable in write access mode.
- **Host Scope Service Engine Within:** SEs may be deployed on any host that most closely matches the resources and reachability criteria for placement. This setting directs the placement of SEs.
    - **Any:** The default setting allows SEs to be deployed to any host that best fits the

deployment criteria.

  - **Cluster:** Excludes SEs from deploying within specified clusters of hosts. Checking the Include checkbox reverses the logic, ensuring SEs only deploy within specified clusters.
  - **Host:** Excludes SEs from deploying on specified hosts. The Include checkbox reverses the logic, ensuring SEs only be deploy within specified hosts.

- **Data Store Scope for Service Engine Virtual Machine:** Set the storage location for SEs. Storage is used to store the OVA (vmdk) file for VMware deployments.
  - **Any:** Vantage will determine the best option for data storage.
  - **Local:** The SE will only use storage on the physical host.
  - **Shared:** Vantage will prefer using the shared storage location. Specific data stores may be Excluded or specified via Include.
- **Virtual Service Placement:** When multiple SE groups exist within a tenant, the virtual service's Advanced tab may be used to choose which SE group to deploy the virtual service within. This may be set as a mandatory field to be populated when creating a virtual service, or when Auto is enabled, the Default-Group will be chosen.
- **Management Network:** If the SEs require a different network for management than the Controller, it must be specified here. The SEs will use their management route to establish communications with the Controllers.

Service Engine Group Network Settings

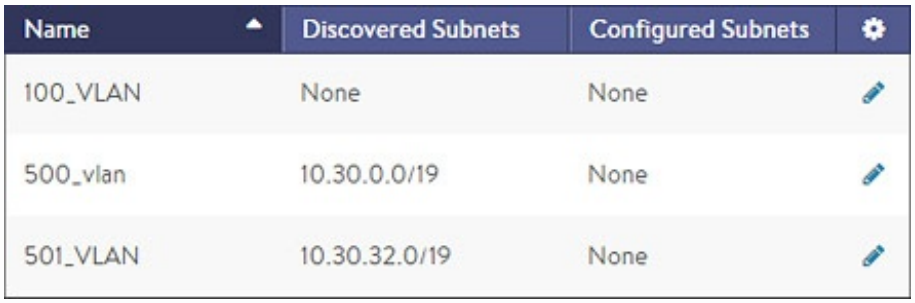
The Networks tab presents the list of discovered and manually configured networks within your network environment. Individual networks can be configured for DHCP or a static IP address allocation. For VMware installations, port groups can be mapped to specific subnets.

- **DVS versus Standard Switching:** VMware supports two modes for switching, Distributed Virtual Switching and Standard Switching. Vantage works with both methods; however, some environments may have both enabled at the same time. This will cause issues for Vantage because there may be multiple port groups per subnet, and the Controller may find duplicate networks for the same IP subnets when performing network discovery. Vantage does not know which network should be used to reach clients or servers and may therefore be unable to place a new virtual service or create a SE in the correct network. You can resolve this by excluding a redundant discovered network. The virtual service Advanced and pool Advanced tabs may alternatively be used to mitigate this issue by mandating a virtual service or pool be placed in a specific network.

- **IP Address Allocation:** Vantage requires IP addresses for a SE to communicate on any desired network. By default, a SE requires one IP address for the management network to communicate with the Controller, and a separate IP address for each data network used by its virtual services or pool servers. If the management network and data network are the same, then the SE will still require two IP addresses. You can allocate IP addresses on either a per-SE basis or via the Networks tab.
- **Network versus Service Engine:** Many network related settings may be configured within both the Network tab and the Service Engine Edit popup. Configurations made within the Network tab will be applied to any new SE created via write access mode. Changes made via the Service Engine Edit popup will only be applied to the specific SE modified.

Select Infrastructure > Networks to open the Networks tab.

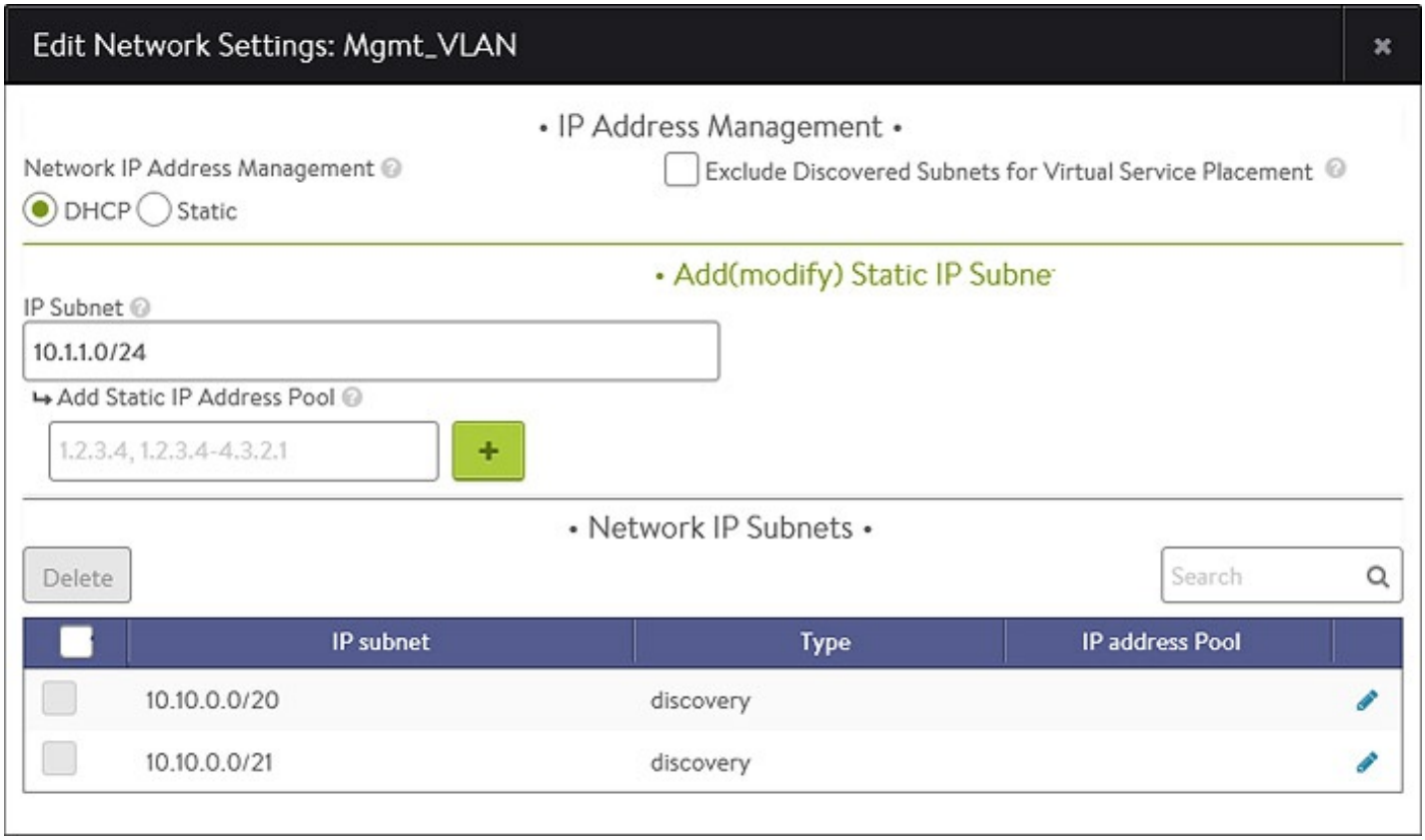
The table on this tab provides the following information for each network:



- **Name:** Name of the network.
- **Discovered Subnets:** These subnets are auto-discovered via the virtualization orchestrator. This field may be None, Excluded, or a list of one or more IP networks.
- **Configured Subnets:** These subnets are IP networks manually added within the Vantage configuration. This is often an IP network that could not be automatically discovered.

Edit Service Engine Group Network Settings

Click the blue Edit icon to open the Edit Network popup.



Enter the following information to edit the network:

- **Network IP Address Management:** When the DHCP option is checked, SEs will attempt to acquire any necessary IP addresses via DHCP. If an SE is unable to acquire an IP address, it will wait five minutes and try again. If no DHCP server is available or if the IP address pool is exhausted, the SE will be unable to properly obtain an IP address and may not be able to configure itself or be able to host a virtual service. Setting this option to Static implies the SE will be assigned static IP addresses.
- **Exclude Discovered Subnets:** IP networks that are discovered in a network or port group will be displayed in the blue table below this option. If there are multiple port groups with the same IP network, Vantage will not know which network should be used for the SEs, Virtual Servers, or when communicating with clients or servers. This is most common for VMware environments that use both DVS and standard switching. Excluding the subnets will exclude all subnets discovered for the network. To exclude a single subnet, first exclude all subnets and then re-add the desired subnets using the Add Subnet option.
- **Add Subnets:** Manually add an IP subnet to this network. Use this options along with Exclude Discovered Subnets to override automated discovery for this network.
    - **IP Subnet:** Specify the IP subnet settings for the new network. For instance: 10.1.1.1/24
    - **Static IP Address Pool:** Instead of using DHCP for IP addresses for this network, SEs can use a statically allocated list of addresses. Add one or more IP addresses, either as

a comma separated list or as a dash-separated range. While possible, it is not recommended to use both DHCP and a static IP pool at the same time. The IP pool allows Vantage to dynamically scale out virtual services and add new SEs. If the IP pool is exhausted for this network, then the Controller may not be able to provision or assign new SEs.

- **Save** to return to the Networks tab.

Static Route

Static routes allow administrators to determine the next hop path for routed traffic. Static routes may be defined for an IP subnet or a specific IP address, determined by the subnet mask defined.

A static route may also be set as the default gateway. Default gateways may also be defined within the settings of an SE, which will override the global static routes, and will be specific to the modified SE. If DHCP is not used and a default gateway needs to be defined, then it is recommended to define the gateway within the Static Routes tab, which will be applicable to all SEs.

Static Routes Tab

Select Infrastructure > Networks > Static Routes to open the Static Routes tab. This tab includes the following functions:

- **Search:** Search through the list of routes.
- **Create:** Opens the Create Static Route popup.
- **Edit:** Opens the Edit Static Route popup.
- **Delete:** Delete the selected static routes.

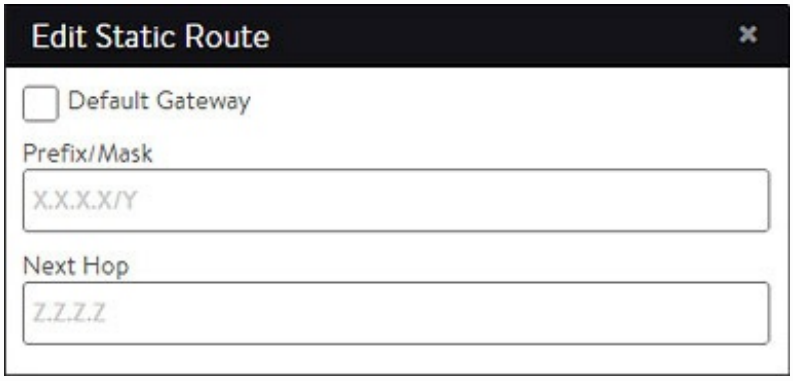
The table on this tab provides the following information for each static route:

- **Index:** Each static route is given a unique identifier, which is used internally for referencing the route.
- **Prefix:** Any egress traffic from Vantage matching this IP subnet will be sent to the IP address of the next hop gateway. A Prefix set to Default Gateway means all traffic that does not match any other static route Prefix will be forwarded to the Next Hop for the default gateway.

- **Next Hop:** The gateway address to use when routing traffic to the IP network specified by the Prefix.

Create/Edit Static Route

The Create Static Route and Edit Static Route popups share the same interface.



Enter the following information to create or edit a static route:

- Check the **Default Gateway** checkbox if this route should be the default for SEs. A default gateway learned from DHCP will override this gateway and will be displayed in an individual SE.
- **Prefix/Mask:** Any egress traffic from Vantage matching this IP subnet will be sent to the IP address of the next hop gateway. A Prefix set to Default Gateway means all traffic that does not match any other Prefix will be forwarded to the Next Hop for this Prefix entry.
- **Next Hop:** The gateway address to use when routing traffic to the IP network specified by the Prefix.
- **Save** to finish adding or editing the static route.

Infrasture

Users

A valid account is required for access to Avi Vantage via the GUI, API or CLI. Avi Vantage maintains these accounts locally or remotely. A user can be:

- Suspended (disabled) or Active.
- Assigned to one or more Tenants.
- Assigned one Role per Tenant which grants permissions within the Tenant.

The Administrator is a unique account that is used for initial setup of Avi Vantage. This account cannot be deleted.

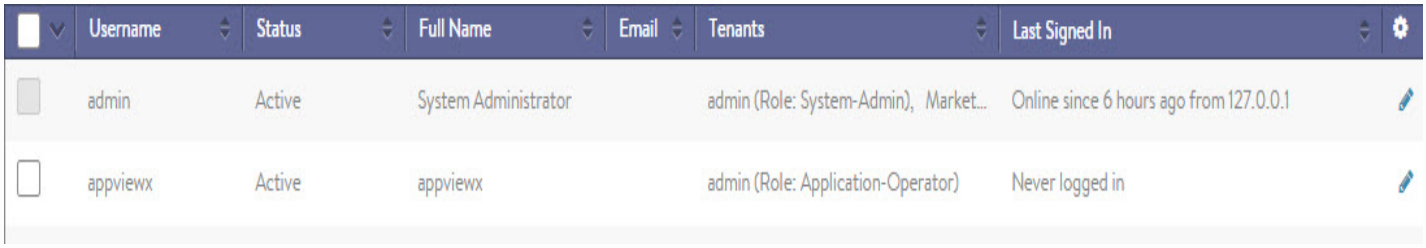
Users Settings

Select Administration > Accounts > Users to open the Users tab.

This tab includes the following functions:

- **Search**
- **Create**
- **Edit**
- **Activate:** Select one or more suspended users and click Activate to reactivate a Suspended user.
- **Suspend:** Select one or more active users and click Suspend to deactivate the selected users. A suspended user cannot access Avi Vantage via the GUI, API, or CLI. That user will receive a notice informing them of the account suspension when they attempt to sign in to Avi Vantage.
- **Delete**

The table on this tab contains the following information for each user:

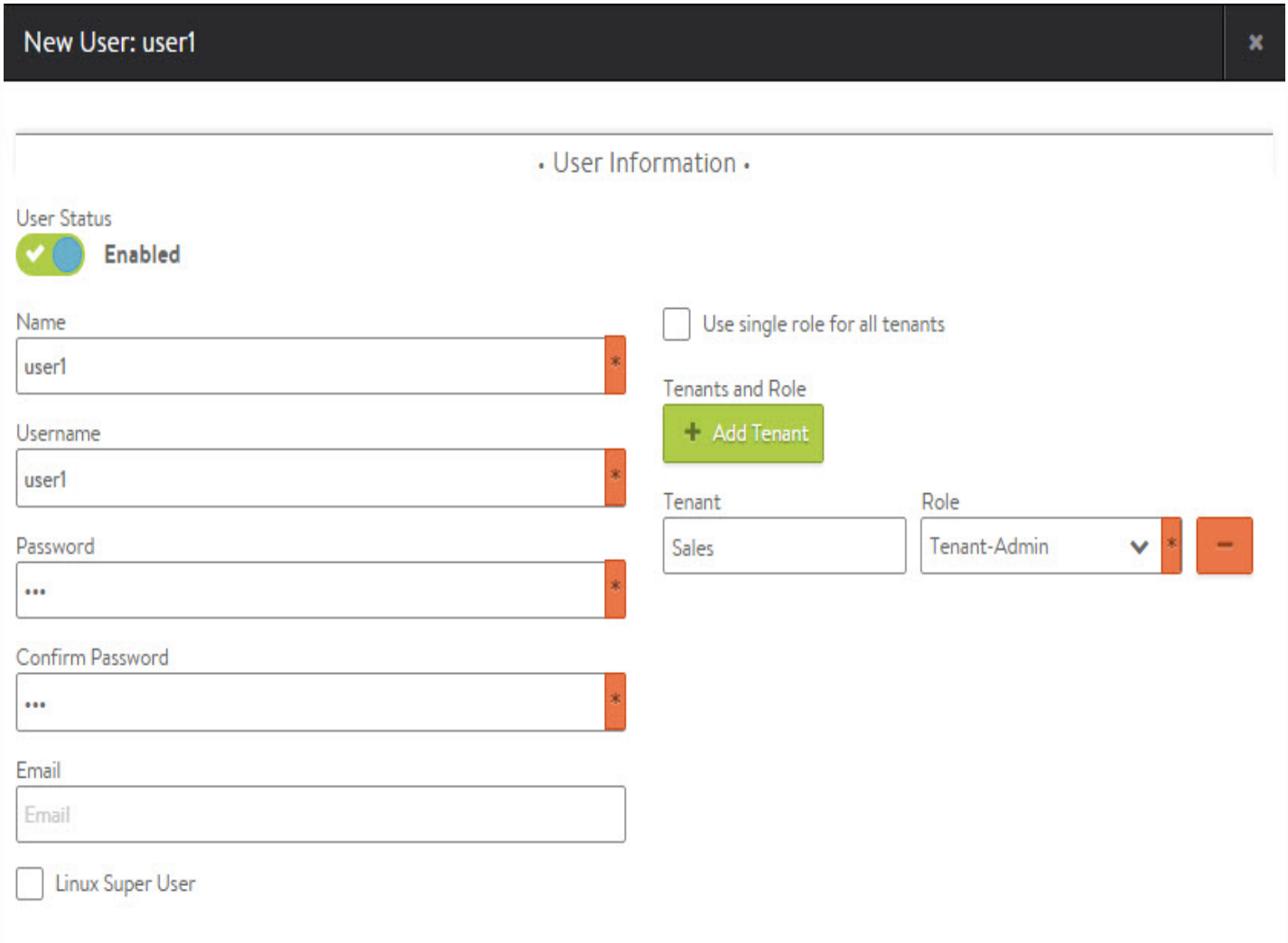


- **User Name:** The account name used to log into Avi Vantage via GUI, API, or CLI.
- **Status:** Status of the user account.
- **Full name:** Full name of the user.
- **Email:** Email address of the user.
- **Tenants:** Tenants the user is allowed to access. (The role assigned to the user defines the access level the user has to specific resources types within the tenant.)

- **Last Signed In:** Date and time that this user last signed in to Avi Vantage via GUI, API, or CLI.

Create a User

The Create User and Edit User popups share the same interface.



Enter the following information to create or edit a role:

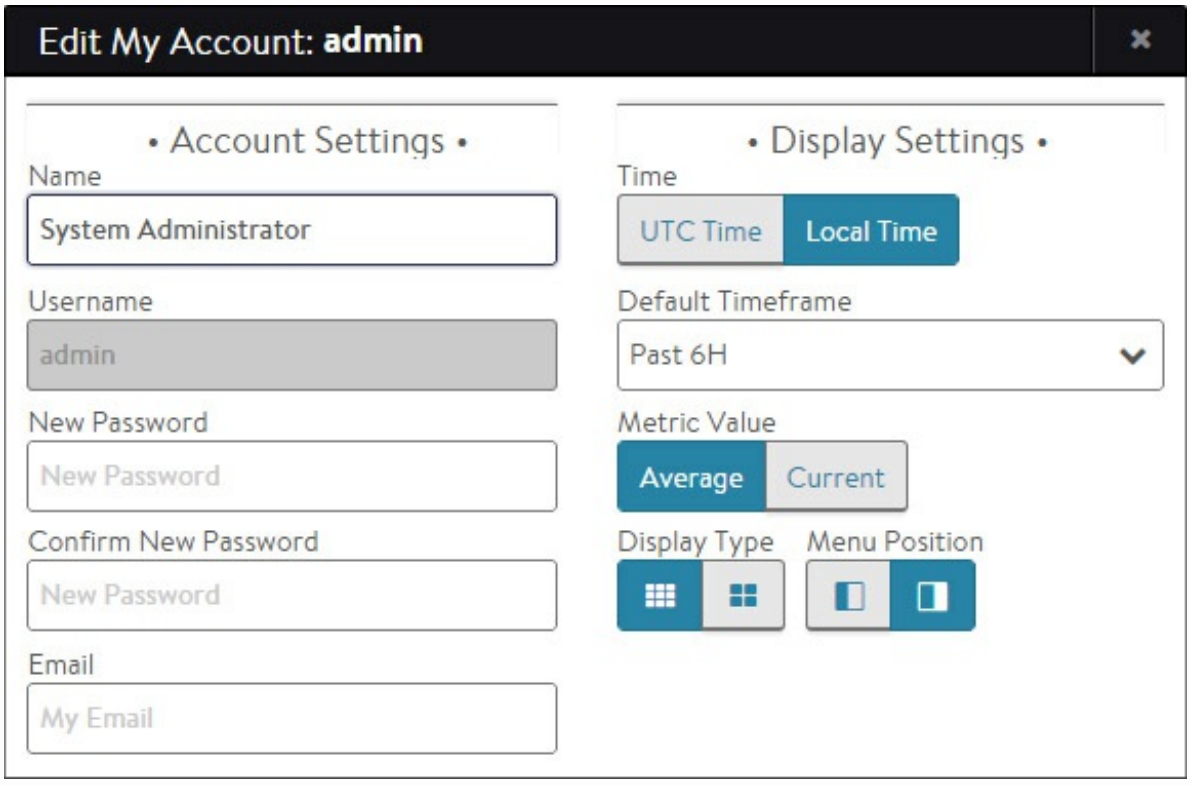
- **Username:** Name that the user will supply when signing into Avi Vantage, such as `jdoe` or `jdoe@avinetworks.com`.
- **Password:** You may either enter a case-sensitive password in this field or click the Generate button to create a random password for the new User.
- **Full Name:** Full name of the user, such as John Doe.
- **Email:** Email address of the user. This field is used when a user loses their password and requests to have it reset. See Password Recovery.
- **Linux Super User:** This checkbox grants root level access to the Linux operating system that

underlies both the Avi Vantage Controller and Service Engines.

- **Tenant:** A user must be assigned a Role within at least one Tenant and may be assigned a Role in each Tenants if you have configured multiple Tenants. Click the Select Tenants button to open the list of configured Tenants, and then select an additional Tenant that you want the user to access. Repeat this for each additional Tenant you want to make accessible to this user.
- **Role:** After selecting each Tenant, you must assign a Role to the user for that Tenant, which determines the level of access that the user has to Avi Vantage for that Tenant. See Roles.
- **Activated/Suspended:** By default, a new user is created in an Active state. The green Enabled switch at the top of the New/Edit User popup allows you to create the user in a Suspended or disabled state by toggling this option off; this switch is green when the user is active, and red when the user is suspended.
- Click **Save** to finish creating or editing the user and return to the Users tab.

User Preferences

A user who is logged into Avi Vantage may edit their account settings by clicking their username in the top right corner of the Avi Vantage UI screen and then selecting My Account to open the Edit My Account popup.



Edit the following default settings for your account:

- **Name:** Your full name.
- **Email:** Your email address. This is the address that Avi Vantage will use when you request a password recovery.
- **Username:** Name assigned to this account and used for signing in to Avi Vantage. This field is display only; you cannot change this setting.
- **Full Name:** The name of the user, such as John Doe.
- **Password:** Enter a new password in the New Password and New Password Again fields (case sensitive).
- Display settings are account-specific preferences that are used each time you sign into Avi Vantage. Changes made to the display settings while navigating the Vantage interface, such as the Display time, will be used for the duration of that session; signing out and back in restores the default preferences stored here. The available options are:
    - **Time:** View timestamps, such as logs or analytics metrics, in either UTC Time (Universal Time Coordinated) or Local Time. Avi Vantage obtains the time from the specified NTP server.
    - **Default Timeframe:** Select the time period over which to display information. See Display Time.
    - **Metric Value:** Change the view of pages such as the Analytics tab of the Virtual Services Details page, to show data points such as the End to End Timing chart as either an Average of the metrics data for the timeframe selected or the most Current data point.
    - **Display Type:** Select either a condensed or expanded view of the Metrics Tiles for tabs that include them.
    - **Menu Position:** Move the Metrics tiles to the right or to the left side of the Analytics tab.
- Save to close the Edit My Account popup and return to the previous page/tab.

Roles

Each User created within Avi Vantage must be assigned a Role, which grants access to resources within Vantage. A User with access to multiple Tenants may have different Roles on a per-Tenant basis. For instance, a User may be given write privileges to the Applications sections of Avi Vantage (virtual services and pools), but only read privileges to all other sections. This access control is maintained for the GUI, API, and CLI.

Creating custom Roles allows you to assign granular Role-Based Access Control (RBAC) within

Avi Vantage. Each user may receive one of the following three access levels to the various areas of the Avi Vantage interface:

- **Write:** User has full access to create, read, modify, and delete items. For example, the user may be able to create a virtual service, modify its properties, view its health and metrics, and later delete that Virtual Server.
- **Read:** User may only read the existing configuration of the item. For example, the user may see how a virtual service is configured while being unable to view the current metrics, modify, or delete that Virtual Server.
- **No Access:** User can neither see nor to modify this section of Avi Vantage. For example, the user would not be able to see the virtual service at all.

Several default roles exist, which provide common combinations of access levels.

- **Application-Admin:** User has Write access to the Application and Profiles sections of Avi Vantage, Read access to the Infrastructure settings, and No Access to the Account or System sections.
- **Application-Operator:** User has Read access to the Application and Profiles sections of Avi Vantage, and No Access to the Infrastructure, Account, and System sections.
- **Security-Admin:** User has Read/Write access only to the Templates > Security section.
- **System-Admin:** User has Write access to all sections of Avi Vantage.
- **Tenant-Admin:** User has Write access to all sections of Avi Vantage except the System section, for which the user has No Access.

Roles Settings

Select Administration > Accounts > Roles to open the Roles tab. This tab includes the following functions:

- **Search**
- **Create**
- **Edit**
- **Delete**

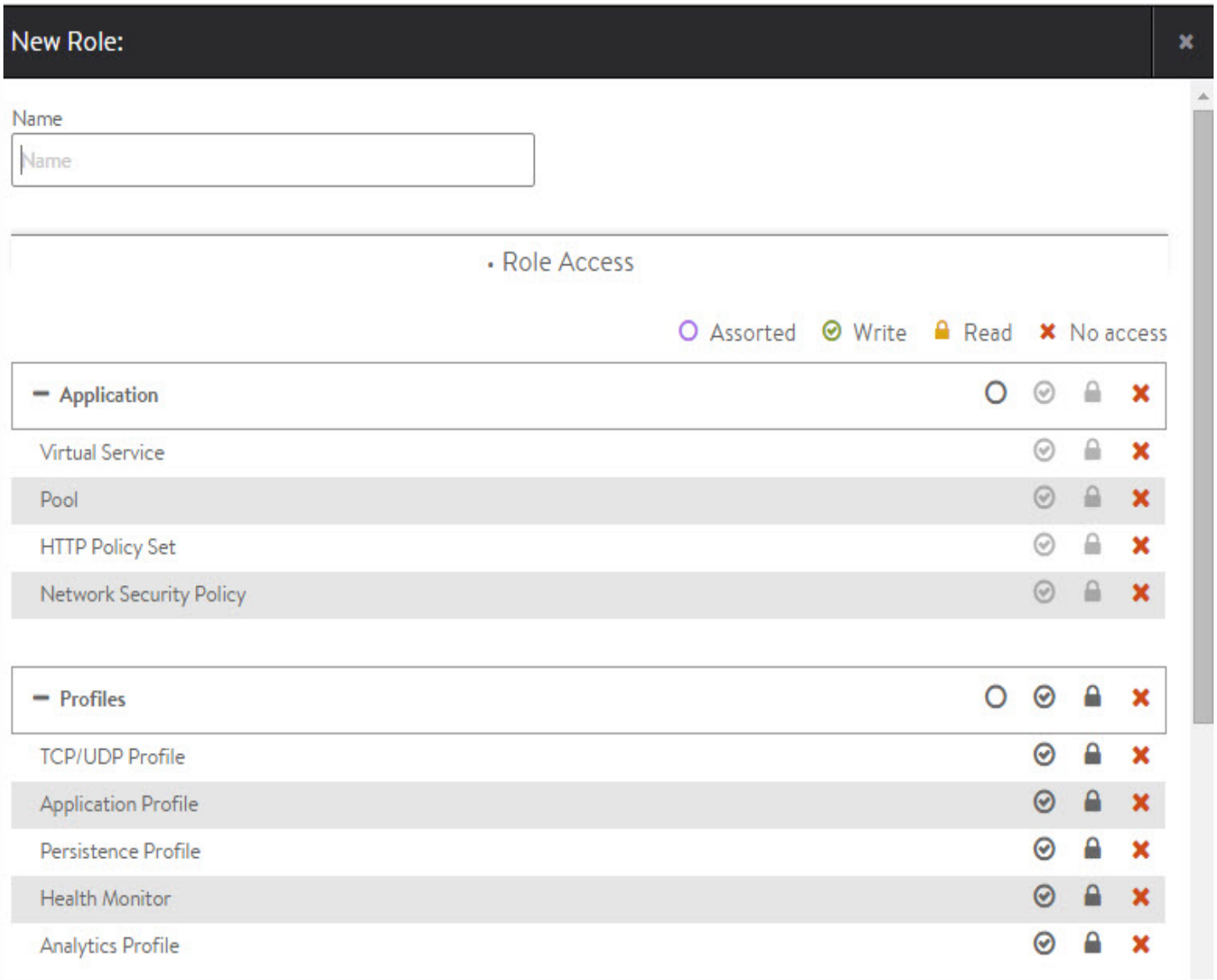
The table on this tab contains the following information for each role:

- **Name:** Name of the Role.

- **Application:** Access level (write, read, or no access) granted to the Role for the Application section of Avi Vantage.
- **Profiles:** Access level granted to the Role for the Profiles section of Avi Vantage.
- **Infrastructure:** Access level granted to the Role for the Infrastructure section of Avi Vantage.
- **Account:** Access level granted to the Role for the Account section of Avi Vantage.
- **System:** Access level granted to the Role for the System section of Avi Vantage.

Create a Role

The Create Role and Edit Role popups share the same interface. To create or edit a Role:



- **Name:** Enter a name in the Role field.
- **Access Level:** Select the desired access level for each of the following sections within Avi Vantage by clicking the appropriate button for each Role (write, read, or no access):

- **Application:** Access level (write, read, or no access) granted to the Role for the Application section of Avi Vantage, which includes Virtual services and pools.
- **Profiles:** Access level granted to the Role for the Profiles section of Avi Vantage.
- **Group & Script:** Access level granted to the Role for the Templates > Groups section of Avi Vantage.
- **Security:** Access level granted to the Role for the Templates > Security section of Avi Vantage.
- **Operations:** Access level granted to the Role for the Operations section of Avi Vantage.
- **Infrastructure:** Access level granted to the Role for the Infrastructure section of Avi Vantage.
- **Administration:** Access level granted to the Role for the Administration of Avi Vantage.
- **Accounts:** Access level granted to the Role for the Account section of Avi Vantage, which includes Users, Roles, and Tenants.

The Assorted icon expands or hides the list of resource types within an access level. When the list is expanded, access can be configured for individual resource types within the access level.

Tenants

A Tenant is an isolated instance of Avi Vantage. An administrator who logs into one Tenant cannot see or view the configuration of other Tenants within Avi Vantage. When working with Tenants:

- If Avi Vantage has been configured with more than one Tenant and if that administrator has been granted access to multiple Tenants, then clicking the username at the top right of the Avi Vantage screen displays a list of available Tenants. Switching to a different Tenant switches the view to show only the configuration for the newly-selected Tenant.

  If Avi Vantage has only been configured with the default Tenant and no others, then all new users and objects (such as pools and virtual services) are assigned to the default Tenant.
- You can create one or more additional Tenants and move both new and existing users into the new Tenants by editing a user account.
- An individual Avi Vantage user may have roles in more than one Tenant; in this case, any new object created by that user will be assigned to the Tenant in which the object was created.

- Service Engine Groups may be created for each Tenant, thereby isolating the data plane resources and preventing contention.

Tenants Settings

Select Administration > Accounts > Tenants to open the Tenants tab. This tab includes the following functions:

- **Search**
- **Create**
- **Edit**
- **Delete**

The table on this tab contains the following information for each Tenant:

- **Name:** Name of the Tenant.
- **Num Users:** Number of users with one or more roles in that Tenant.
- **Admin Users:** Usernames of users with administrator privileges in a Tenant.
- **Other Users:** Usernames of users with non-administrator privileges in a Tenant.

Create a Tenant

The Create Tenant and Edit Tenant popups share the same interface.

To create or edit a Tenant:

- **Name:** Enter a Name for the new Tenant.
- **Save** to finish creating the new Tenant and return to the Tenants tab.

Once the Tenant has been created, users can be moved into or created in the Tenant.

Settings

Access Settings Tab

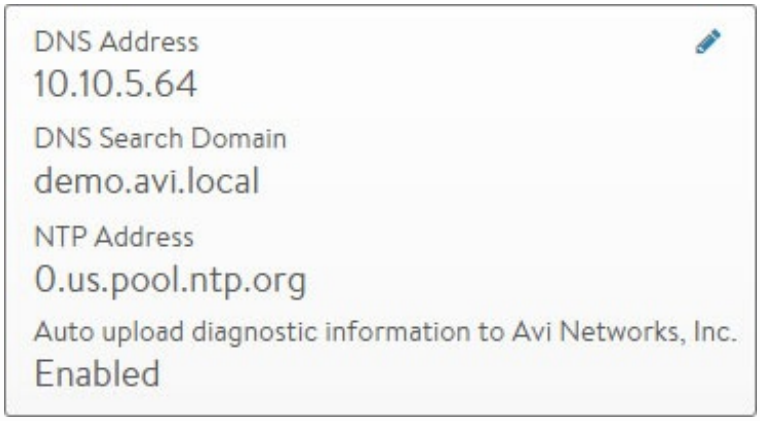
Select Administration > Settings > Access Settings to open the System Access tab.

These options govern how administrators may access the Avi Controller UI.

- **HTTPS Access to System:** Enables SSL/TLS access to the Avi Vantage's GUI and REST API. When this option is enabled, the SSL Profile and SSL/TLS Certificate fields must be populated.
- **HTTP Access to System:** Allow HTTP access to the Avi Cloud ADC's GUI and REST API. This option is insecure and not recommended.
- **Redirect HTTP to HTTPS:** When HTTP Access to System is disabled, enabling this option will automatically redirect administrators to the HTTPS interface for the GUI and API.
- **SSL Profile:** Select an SSL Profile to complete the HTTPS Access. This profile is from the Templates > Security > SSL Profiles, which is also referenced by SSL enabled Virtual Services.
- **SSL/TLS Certificate:** Select an SSL certificate to present to clients connecting to the Avi Vantage GUI. This may be either an RSA or Elliptic Curve certificate.

DNS / NTP Settings

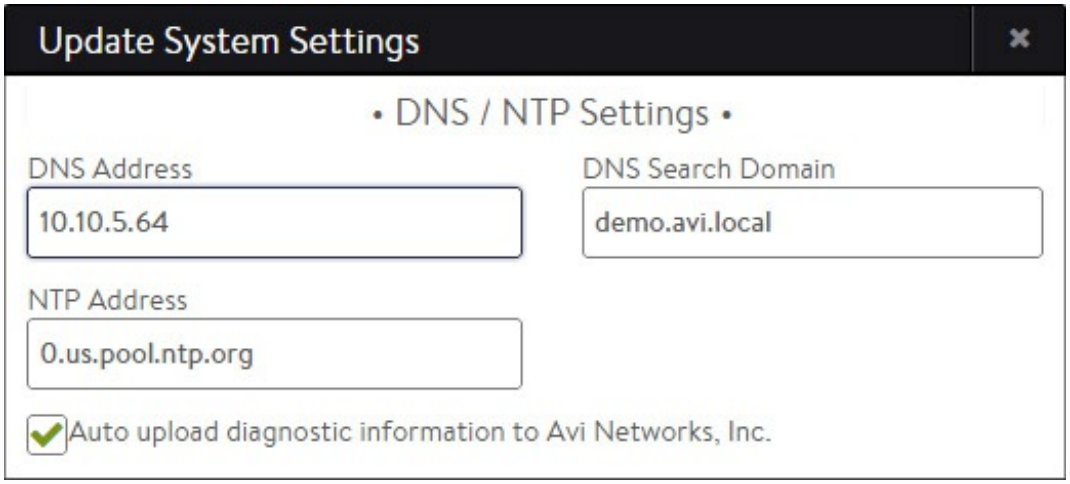
Select Administration > Settings > DNS/NTP to open the DNS/NTP tab of the Settings page. This tab presents the following information:



- **DNS Address:** IP address of the domain name server.
- **DNS Search Domain:** Local domain name, which will be appended to a name that is not fully qualified. For instance, if the DNS Search Domain is set to avinetworks.com, and the name to be resolved is www, then Avi Vantage will lookup www.avinetworks.com.
- **NTP Address:** The IP address of a Network Time Protocol server that Avi Vantage will use to maintain date and time. NTP will be used to set the clock on the Controllers. The Controllers will then set the clock on the Service Engines. If no NTP is configured or accessible, the Controllers clock may drift or be off, but the Service Engines will still be in sync with the Controllers time.
- **Auto upload diagnostic information to Avi Networks:** When selected, Avi Vantage will automatically send utilization data to Avi Networks. This information contains crash reports and system logs which Avi may use to diagnose issues. It does not contain sensitive information, such as passwords or SSL certificates. For more information on the exact data sent to Avi Networks, contact Avi's technical support.

Update DNS/NTP Settings

The blue Edit icon on the DNS/NTP tab of the Settings page opens the Update System Settings popup, which configures the basic network settings used by Avi Vantage. These settings were initially defined during the first time setup.



Enter the following information to update the DNS/NTP settings:

- **DNS Address:** Enter the IP address of the Domain Name Server. When a name server is not configured, Avi Vantage will not be able to accept names for load balanced servers, Virtual Services, mail servers, and similar inputs.
- **DNS Search Domain:** Local domain name (optional). This will be appended to a name that is not fully qualified. For example, if the DNS Search Domain is set to avinetworks.com, and the name to be resolved is www, Avi Vantage will lookup www.avinetworks.com.
- **NTP Address:** IP address or hostname of the Network Time Protocol servers that Avi Vantage will use to obtain the correct local time. If no NTP server is defined, Avi Vantage will use the time from the hypervisor used by the Controller. Avi Vantage Controllers and Service Engines will remain in sync with each other, but not necessarily with the outside world. This may affect time stamps for logs or other areas where times may need to be correlated with external devices for troubleshooting.
- **Auto upload diagnostic information to Avi Networks:**Avi Vantage will automatically send diagnostic and troubleshooting information to Avi Networks, such as crash reports. For instance, if Avi Vantage experiences a crash, a report would be sent to Avi Networks with details about the issue. Sensitive information such as passwords and SSL certificates will not be transmitted. Clearing this checkbox disables automatic uploads. Contact Avi Network's technical support for current details about the exact data that is captured.

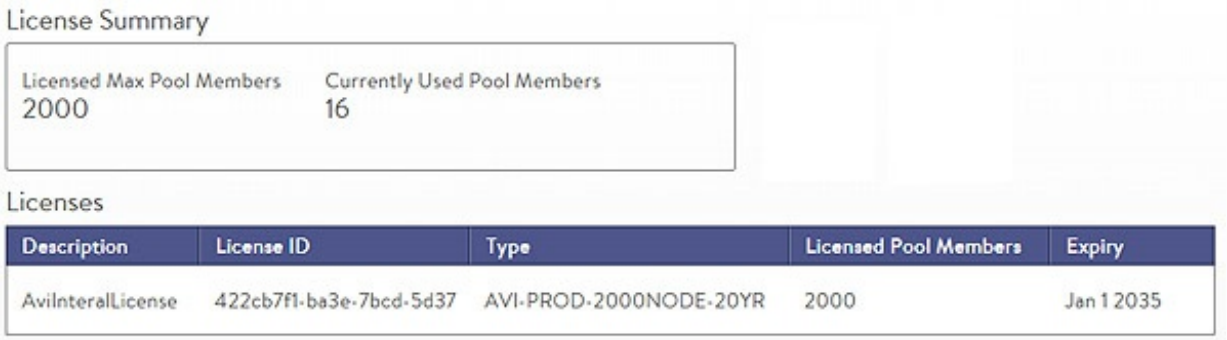
License Tab

Select Administration > Settings > Licensing to open the Licensing tab.

Avi Vantage is bound by a license that determines the features and capacity available to a Controller cluster installation. When a valid license file is not available, Avi Vantage runs in Free Trial mode, which limits the total system capacity. In this mode, Vantage may only load balance two servers within a pool, and is limited to two Service Engines. All functionality, such as SSL or

DataScripts are still available.

A new or additional license may be added at any time to provide increased capacity. Once a license has been added to a Controller, it has been consumed and cannot be removed.



This tab presents the following information:

- **License Summary:**
  - **Licensed Max Pool Members:** Avi Cloud ADC is metered by the maximum number of servers that may be configured within Pools. This number is the total number of all Pool members at any time. The max shows how many Pool members the system is licensed to support. This number may be cumulative from multiple licenses.
  - **Currently Used Pool Members:** This stat shows the number of Pool members currently configured across the system. If more servers are added to a Pool than the license will allow, the process of saving the change will fail and an error will be generated. As servers are removed from Pool or Pools are deleted, the current Pool member's number will drop.
- **Licenses:**
  - **Description:** The description field denotes the company or organization that owns the license.
  - **License ID:** The unique serial number for the license file.
  - **Type:** The specific license type, such as production or trial. This entry also may state the number of nodes, or Pool members and the length of time the license is valid.
  - **Licensed Pool Members:** The number of Pool members this license will add to the Licensed Max Pool Members stat.
  - **Expiry:** The length of time this license is considered valid. If all licenses have expired, Avi Vantage will revert to a trial mode, which allows all features but limits the configuration to supporting the minimum number of Pool members.

Click the green Upload License File button to open a popup window that allows you to navigate to, select, and upload a new license file to activate, reactivate, or modify your Avi Vantage license.

Email/SMTP Tab

Select Administration > Settings > Email/SMTP to open the Email/SMTP tab.

Avi Vantage may proactively send emails. This may be done for password reset operations, or to send defined Alert Actions to administrators or monitoring systems. Emails are sent from the Avi Controller, which means the Controller will require DNS and network access to a destination email server.

The SMTP Source is where the email will be source from when the Avi Controller is sending an email.

- **None:** When set to None, Avi Vantage will not send emails.
- **Local Host:** The Avi Controller will send the email from a local host. Some enterprise email servers may not accept this method.
    - **From Address:** The email field that denotes who sent the email.
- **SMTP Server:** Avi Vantage can point to a remote SMTP server to use for sending emails. This method is generally more trusted by security conscious enterprise environments.
    - **Username:** The username to authenticate to the mail server.
    - **Password:** The password to authenticate to the mail server.
    - **SMTP Server:** The mail server host.
    - **Port Number:** The service port for SMTP, generally 25.
    - **From Address:** The email field that denotes who sent the email.

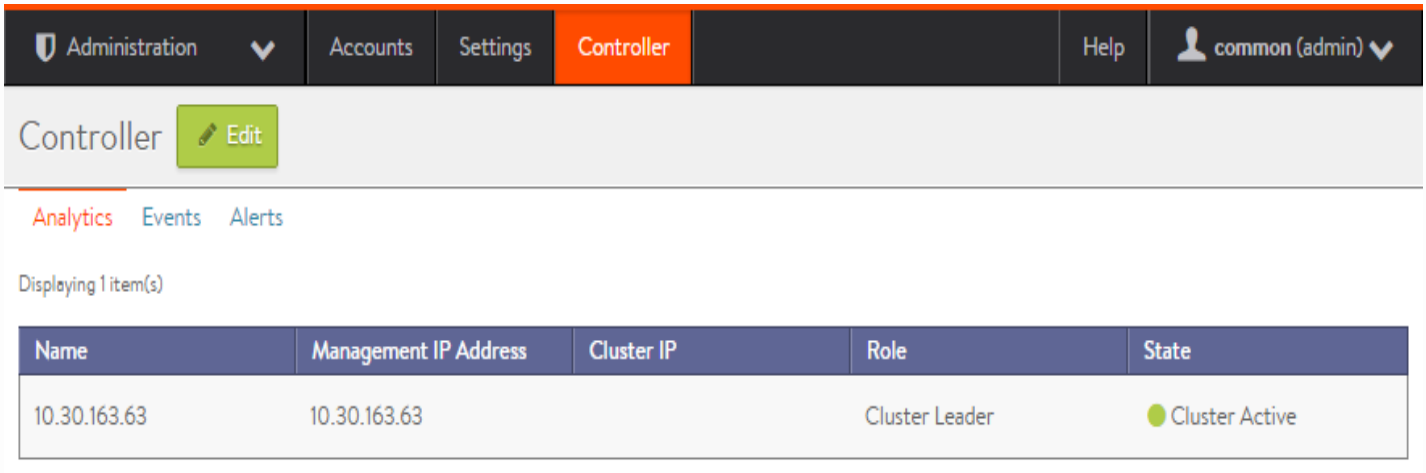
Controller

The Administration > Controller page provides information about the Controller (or Controller cluster).

Controller Table

The Controller table lists the Controllers. If Vantage is deployed with a single Controller, only that

Controller is listed. If Vantage is deployed as a 3-node Controller cluster, each Controller node is listed. The following example is from a deployment that uses a single Controller:



The Controller table contains the following information:

- **Name:** Name or IP address of the Controller.
- **Management IP Address:** Management IP address of the individual Controller (node) in the cluster. Each Controller in the cluster will have its own management address.
- **Cluster IP:** The shared management IP address for the cluster, owned by the current primary (leader) Controller.
- **Role:** Shows which Controller is the primary, or leader of the cluster.
- **State:** The Controller's status, such as "Cluster Active".

Controller Analytics

This tab configures the management IP addresses for the Controller(s).

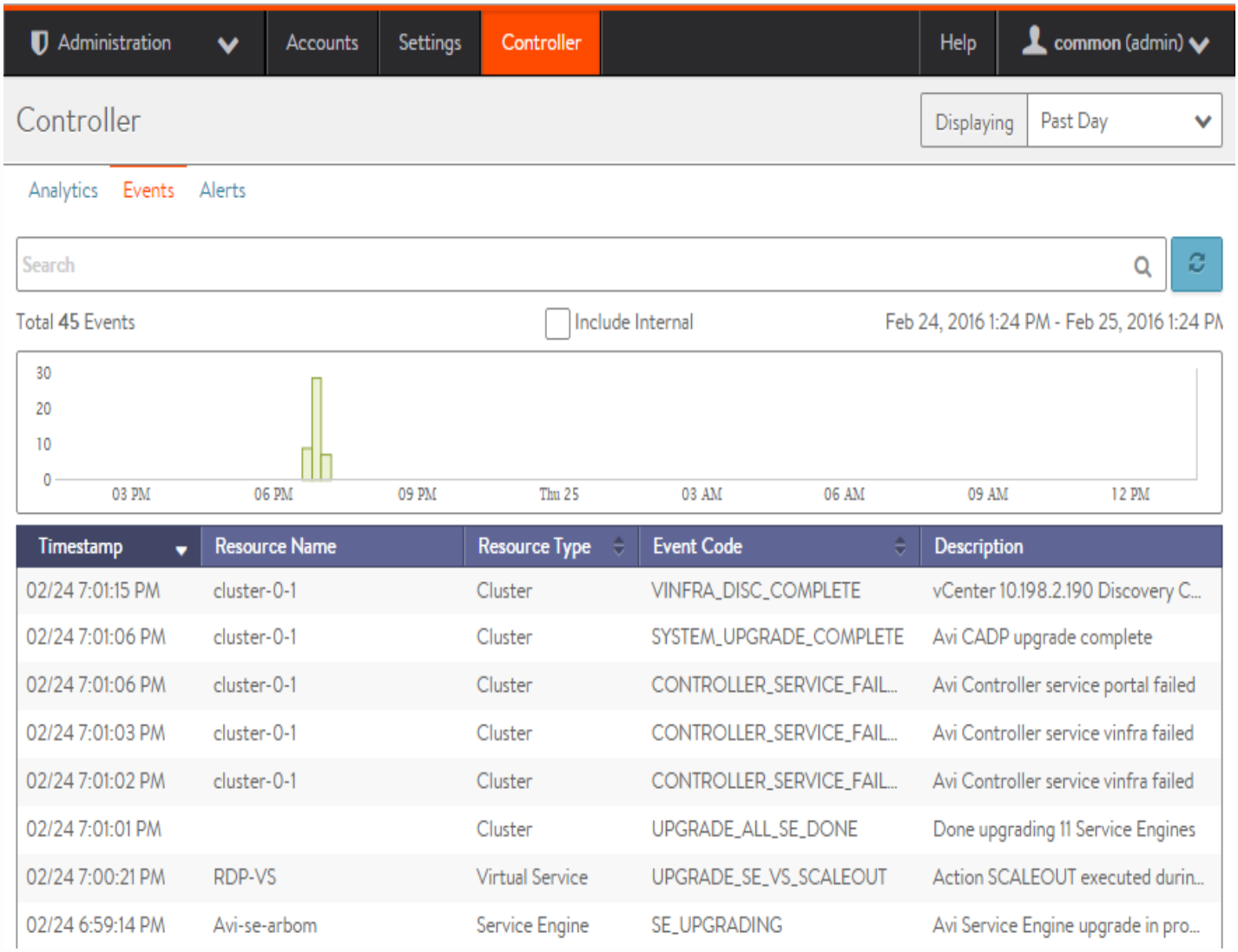
In a 3-node cluster, each Controller node will have its own management IP address, which can be directly accessed. The cluster will also have a single address that moves to whichever Controller is currently the cluster leader.

Best practice is to always access the Controller cluster through the cluster IP address, in the event there is an issue with a single Controller. By using the cluster IP address for access, access to the Controller web interface can be maintained even while an individual Controller is offline. For example, the cluster IP address provides consistent management access to Vantage during the upgrade process.

- **Cluster Information:**
  - **Controller Cluster IP:** The shared management IP address for the cluster.
- **Management IP:**
  - **Controller Node-1:** Management IP address of node 1. This IP interface belongs only to this Controller node and is not shared by the cluster.
  - **Controller Node-2:** Management IP address of node 2. This IP interface belongs only to this Controller node and is not shared by the cluster.
  - **Controller Node-3:** Management IP address of node 3. This IP interface belongs only to this Controller node and is not shared by the cluster.

Controller Events

This page lists Controller-related events. This page automatically filters the Vantage events log to display only those events related to Controller operation. The same events also are included in the events log accessed by Operations > Events. The same search options also are supported.



Controller Alerts

This page lists Controller-related alerts. This page automatically filters the Vantage alerts log to display only those alerts related to Controller operation. The same alerts also are included in the alerts log accessed by Operations > Alerts. The same search options also are supported.

:etoN . desugaftaV ehtyb

Converting a Single-node Cluster to a Three-node Cluster

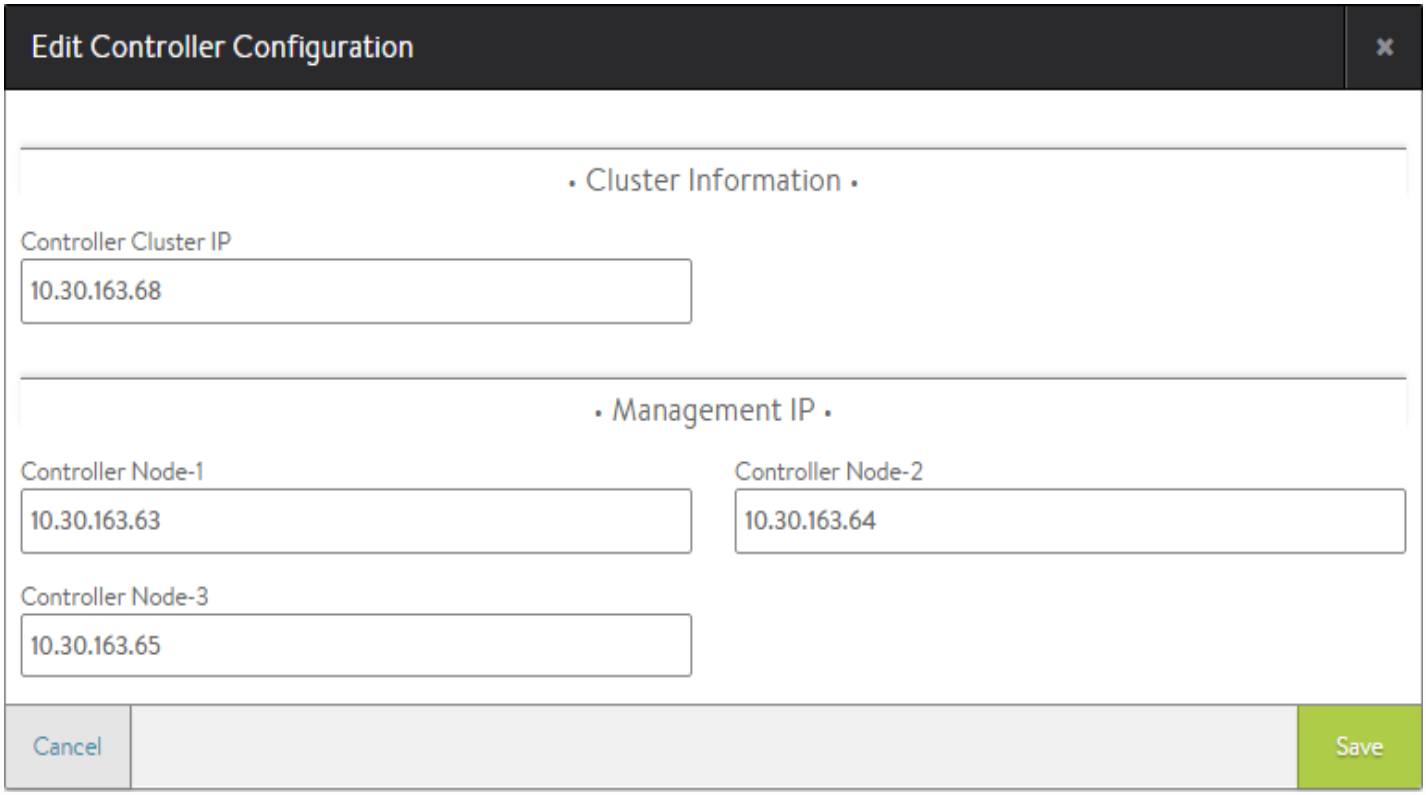
If Vantage was deployed with a single Controller node, it is possible to add 2 additional Controller nodes to form a 3-node cluster. To ensure complete system redundancy, the Avi Controller must be highly available. High availability of the Controller requires 3 separate Controller instances,

configured as a 3-node cluster.

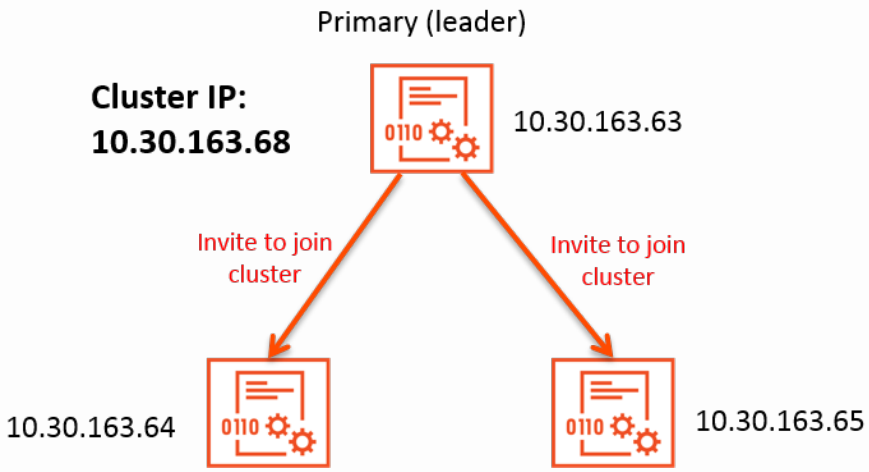
To convert a single-node Controller deployment into a 3-node deployment, use the following steps.

In this procedure, the Controller node that is already deployed in the singe-node deployment is referred to as the *incumbent* Controller.

1. Install 2 (two) new Controller nodes. During installation, configure ***only*** the following settings for each node:
    - Node management IP address
    - Gateway address
2. Connect the management interface of each new Controller node to the same network as the incumbent Controller. After the incumbent Controller detects the 2 new Controller nodes, the incumbent Controller will become the primary (leader) Controller for the 3-node cluster.
3. Use a web browser to navigate to the management IP address of the primary (leader) Controller.
4. Navigate to Administrator > Controller, and click Edit. The Edit Controller Configuration popup appears.
5. In the Controller Cluster IP field, enter the shared IP address for the Controller cluster.
6. In the Controller Node-2 and Controller Node-3 fields, enter the management IP addresses of the new Controller nodes.



After these steps, the incumbent Controller becomes the primary (leader) for the cluster and invites the other Controllers to the cluster as members. Vantage then performs a warm reboot of the cluster. This process can take 2-3 minutes. The configuration of the primary (leader) Controller is synchronized to the new member nodes when the cluster comes online following the reboot.



High Availability Operation for Controllers

Vantage provides high availability for virtual services, which is separately configurable within individual virtual services (using the Advanced tab of a virtual service's configuration.)

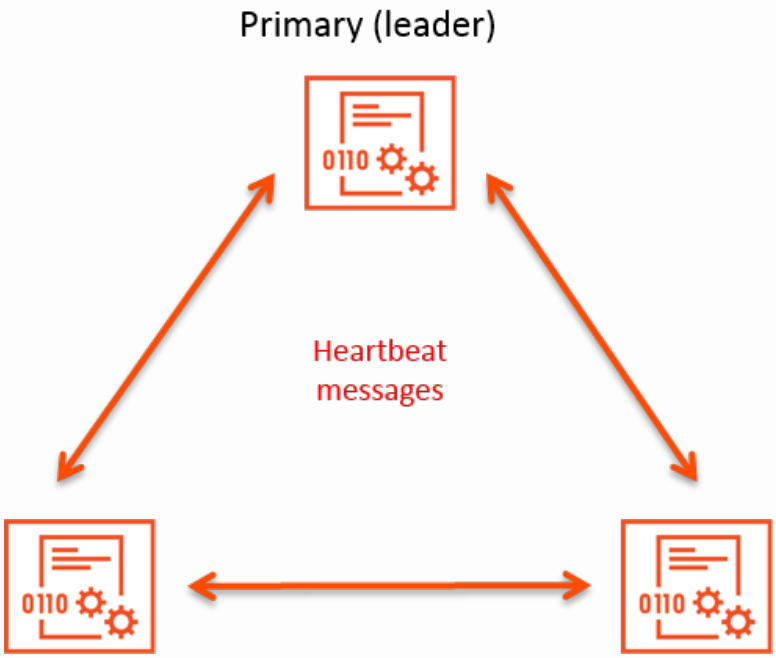
In addition to high availability for individual virtual services, the Controllers in a 3-node cluster provide an additional level of high availability with redundancy of the Controller itself.

Quorum

Controller-level high availability requires a quorum of Controller nodes to be up. In a 3-node Controller cluster, quorum can be maintained if at least 2 of the 3 Controller nodes are up. If one of the Controllers fails, the remaining 2 nodes continue service and Vantage continues to operate. However, if 2 of the 3 nodes go down, the entire cluster goes down and Vantage stops working.

Failover

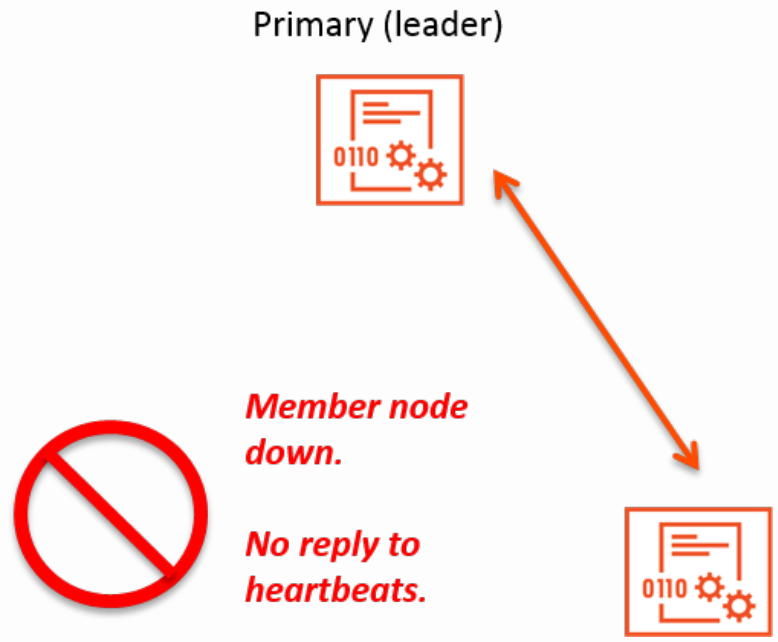
Each Controller node in a cluster periodically sends heartbeat messages to the other Controller nodes in the cluster.



The heartbeat interval is 10 seconds. The maximum number of consecutive heartbeat messages that can be missed is 4. If a Controller does not hear from another Controller for 40 seconds (4 missed heartbeats), the other Controller is assumed to be down.

If only one node is down, quorum is maintained and the cluster can continue to operate.

- If a member node goes down by the primary (leader) node stays, access to virtual services continues without interruption.



- If the primary (leader) node goes down, the member nodes form a new quorum and elect a cluster leader. The election process takes about 50-60 seconds and during this period, there is no impact on the data plane. The SEs will continue to operate in “headless mode” but control plane service is not going to be available. During this period, Vantage users will not be able to create a VIP through LBaaS or using the Vantage web interface, API, or CLI.