



Service Engine Group

Avi Technical Reference (v18.1)

Copyright © 2020

Service Engine Group

[view online](#)

Overview

Service Engines are created within a group, which contains the definition of how the SEs should be sized, placed, and made highly available. Each [cloud](#) will have at least one SE group. The options within an SE group may vary based on the type of cloud within which they exist and its settings, such as [no access](#) versus [write access](#) mode. A given SE may only exist within one group. Each group acts as an isolation domain. SE resources within an SE group may be moved around to accommodate virtual services, but SE resources are never shared between SE groups.

Depending on the change, a change made to an SE group * may be applied immediately, * only applied to SEs created after the changes are made, or * require existing SEs to first be [disabled](#) before the changes can take effect.

Multiple SE groups may exist within a cloud. A newly created virtual service will be placed on the default SE group, though this can be changed via the VS > Advanced page while creating a VS. To move an existing virtual service from one SE group to another, the VS must first be disabled, moved, and then re-enabled. SE groups provide data-plane isolation. Consequently, moving a VS from one SE group to another is disruptive to existing connections through the virtual service.

The screenshot shows the 'Default-Cloud' configuration page in the Avi Networks Infrastructure console. The 'Service Engine Group' tab is selected. A green button labeled '+ CREATE SERVICE ENGINE GROUP' is visible. Below it is a search bar and a table with the following data:

Name	# Service Engines	Maximum Number of...	# Virtual Services	HA Mode
Default-Group	1	10	8	N+M (buffer)
SE group 1	0	2	1	Legacy Active/Standby

Note ? In some clouds, not all settings described below are available.

SE Group ? Basic Settings Tab

To access the Service Engine group editor, navigate to Infrastructure > Service Engine Group. Click the pencil icon to edit a pre-existing SE group, or click the green button to create a new one. Start your specification for the SE group by giving it a name (or accept the default name, which is *Default-Group*), as shown below.

Real-Time Metrics

The screenshot shows a configuration window titled "New Service Engine Group:". It has two tabs: "Basic Settings" (selected) and "Advanced". Under "Basic Settings", there is a text input field for "Service Engine Group Name*" with the placeholder "Name". To the right, there is a "Metric Update Frequency" section with a help icon. It contains a checked checkbox for "Real-Time Metrics", a numeric input field with the value "30", and a "min" label.

At the top right of the Basic Settings tab, check the box to turn on real-time metrics, which will cause SEs in the group to upload SE-related metrics to the Controller once every 5 seconds, as opposed to once per five minutes or longer. [[More info on metrics-upload intervals.](#)] After clicking the box, select the number of minutes for real-time updating to last. The default is 30 minutes; a value of 0 is interpreted to mean "forever."

High Availability & Placement Settings

The high availability mode of the SE group controls the behavior of the SE group in the event of an SE failure. It also controls how load is scaled across SEs. Selecting a particular HA mode will change the settings and options that are exposed in the UI. These modes span a spectrum, from use of the fewest virtual machine resources on one end to providing the best high availability on the other.

- [Legacy HA Active Standby Mode](#): This mode is primarily intended to mimic a legacy appliance load balancer for easy migration to Avi Vantage. Only two Service Engines may be created. For every virtual service active on one, there is a standby on the other, configured and ready to take over in case of a failure of the active SE. There is no Service Engine scale out in this HA mode.
- [Elastic HA N + M HA](#): This default mode permits up to N active SEs to deliver virtual services, with the capacity equivalent of M SEs within the group ready to absorb SE(s) failure(s).
- [Elastic HA Active/Active HA](#): This HA mode distributes virtual services across a minimum of two SEs.

For additional considerations for SE high availability, including VS placement, see [Overview of Avi Vantage High Availability](#).

- **VS Placement across SEs ?** When placement is *compact* (previously referred to as "Compactor"), Avi Vantage prefers to spin up and fill up the minimum number of SEs; it tries to place virtual services on SEs which are already running. When placement is *distributed*, Avi Vantage maximizes VS performance by avoiding placements on existing SEs. Instead, it places virtual services on newly spun-up SEs, up to Max Number of Service Engines. By default, placement is compact for elastic HA N+M mode and legacy HA active/standby mode. By default, it is distributed for elastic HA active/active mode.
- **Virtual Services per Service Engine ?** This parameter establishes the maximum number of virtual services the Controller cluster can place on any one of the SEs in the group.
- **Per Application SE mode ?** Select this option to deploy dedicated load balancers per application, i.e., per virtual service. In this mode, each SE is limited to a maximum of 2 virtual services. vCPUs in per-app SEs count towards licensing at 25% rate.
- **SE Self-Election ?** Checking this option enables SEs in the group to elect a primary SE amongst themselves in the absence of connectivity to a Controller. This ensures Service Engine high availability in handling client traffic even in headless mode. For more information, refer to the [Service Engine Self Election](#) article.

Service Engine Capacity & Limit Settings

• Service Engine Capacity and Limit Settings •

<p>Max Number of Service Engines ?</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 10 Maximum </div>	<p>Memory per Service Engine ?</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 2048 MB </div>	<p>vCPU per Service Engine ?</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 2 </div>	<p>Disk per Service Engine ?</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 10 GB </div>
<p><input checked="" type="checkbox"/> Memory Reserve <input type="checkbox"/> CPU Reserve</p>			
<p>Connection Memory Percentage (slide the bar left or right) ?</p> <div style="border: 1px solid #ccc; padding: 5px; position: relative; width: 100%;"> <div style="background-color: #92d050; width: 50%; height: 20px; position: absolute; top: -20px; left: -20px;"></div> </div>			
Connections: 50%		Buffers: 50%	

- **Max Number of Service Engines ?** [Default = 10, range = 0-1000] Defines the maximum SEs that may be created within an SE group. This number, combined with the virtual services per SE setting, dictates 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, un-deployed status. This setting can be useful to prevent Avi Vantage from consuming too many virtual machines.
- **Memory per Service Engine ?** [Default = 2 GB, min = 1 GB] Enter the amount of RAM, in multiples of 1024 MB, to allocate to all new SEs. Changes to this field will only affect newly-created 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 ?** [Default is ON] 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 strongly 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 ?** [Default = 1, range = 1-64] 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 ?** [Default is OFF] Reserving CPU capacity with a virtualization orchestrator ensures an 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.
- **Disk per Service Engine ?** [min = 10 GB] Specify an integral number of GB of disk to allocate to all new SEs. This option is only applicable in write access mode deployments. The value appearing in the window is either:
 - 10 GB (the absolute minimum allowed), or
 - a value auto-calculated by the UI as follows: 5 GB + 2 x memory-per-SE, or
 - a number explicitly keyed in by the user (values less than 5 GB + 2 x memory-per-SE will be rejected)
- **Host Geo Profile ?** [Default is OFF] Enabling this provides extra configuration memory to support a large geo DB configuration.
- **Connection Memory Percentage ?** The percentage of memory reserved to maintain connection state. It comes at the expense of memory used for HTTP in-memory cache. Sliding the bar causes the percentage devoted to connection state to range between its limits, 10% minimum and 90% maximum.

License

• License •

Pay-as-you-go
 Bring your own license

License Type ? Cores
 License Tier ? Enterprise 18

Allow Burst ?

- License Tier ? Specifies the license tier to be used by new SE groups. By default, this field inherits the value from the system configuration.
- License Type ? If no license type is specified, Avi applies default license enforcement for the cloud type. The default mappings are max SEs for a container cloud, cores for OpenStack and VMware, and sockets for Linux.
- SE Bandwith Type ? Select the SE bandwidth for the bandwidth license.

SE Group ? Advanced Tab

The advanced tab in the Service Engine group popup supports configuration of optional functionality for SE groups. This tab only exists for clouds configured with write access mode. The appearance of some fields is contingent upon selections made.

Basic Settings
Advanced

Service Engine Name Prefix ? Avi
 Delete Unused Service Engines After ? 10 Min

Service Engine Folder ? AviSeFolder

• Host & Data Store Scope •

Host Scope Service Engine within: Any Cluster Host

Data Store Scope for Service Engine Virtual Machine: Any Local Shared

• Advanced HA & Placement •

Buffer Service Engines ? 0
 Scale per Virtual Service ? 1 Minimum 4 Maximum

Override Management Network ? Select a Network

CPU socket Affinity ?
 Dedicated dispatcher CPU ?

• Security •

HSM Group Select HSM Group

• Log Throttle Settings •

Significant Log Throttle ? 100 Logs/Second
 UDF Log Throttle ? 100 Logs/Second
 Non-significant Log Throttle ? 100 Logs/Second

- **Service Engine Name Prefix ?** Enter the prefix to use when naming the SEs within the SE group. This name will be seen both within Avi 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 Avi 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.

Host & Data Store Scope

- **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 ?** Sets the storage location for SEs. Storage is used to store the OVA (vmdk) file for VMware deployments.

Data Store Scope for Service Engine Virtual Machine

Data Store

Include
 Exclude

- **Any ?** Avi Vantage will determine the best option for data storage.
- **Local ?** The SE will only use storage on the physical host.
- **Shared ?** Avi Vantage will prefer using the shared storage location. When this option is clicked, specific data stores may be identified for exclusion or inclusion.

Advanced HA & Placement

- **Buffer Service Engines ?** This is excess capacity provisioned for HA failover. In elastic HA N+M mode, this is capacity is expressed as M, an integer number of buffer service engines. It actually translates into a count of potential VS placements.



To calculate that count, Avi Vantage multiplies M by the maximum number of virtual services per SE. For example, if one requests 2 buffer SEs (M=2) and the max_VS_per_SE is 5, the count is 10. If max SEs/group hasn't been reached,

Avi Vantage will spin up additional SEs to maintain the ability to perform 10 placements. As illustrated at right, six virtual services have already been placed, and the current count of spare capacity is 14, more than enough to perform 10 placements. When SE2 fills up, spare capacity will be just right. An 11th placement on SE3 would reduce the count to 9 and require SE5 to be spun up.

- **Scale Per Virtual Service ?** A pair of integers determine the minimum and maximum number of active SEs any virtual service within this group can scale out to. With native SE scaling, the greatest value one can enter as a maximum is 4; with BGP-based SE scaling, the limit is much higher, governed by the ECMP support on the upstream router.
- **See also:**
[BGP Support for Scaling Virtual Services](#)
[Impact of Changes to Min/Max Scaleout per Virtual Service.](#)
- **Override 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. For more information, read [Deploy SEs in Different Datacenter from Controllers.](#)

Note ? This option is only available if the SE group's overridden management network is DHCP-defined. An administrator's attempt to override a statically-defined management network (Infrastructure > Cloud > Network) will not work due to not allowing a default gateway in the statically-defined subnet.

- **CPU socket Affinity ?** Selecting this option causes Avi Vantage to allocate all cores for SE VMs on the same socket of a multi-socket CPU. The option is applicable only in vCenter environments. Appropriate physical resources need to be present in the ESX Host. If not, then SE creation will fail and manual intervention will be required.

Note: The vCenter drop-down list populates the datastores if the datastores are shared. The non-shared datastores (which means each ESX Host has their own local datastore) are filtered out from the list because, by default when an ESX Host is chosen for SE VM creation, the local datastore of that ESX Host will be picked.

- **Dedicated dispatcher CPU:** Selecting this option dedicates the core that handles packet receive/transmit from/to the data network to just the dispatching function. This option makes most sense in a group whose SEs have three or more vCPUs.

Self Service Engine Election

Starting with Avi Vantage release 18.1.2, in the absence of Controller connectivity, Service Engines can elect a primary SE within the HA group. This ensures Service Engine high availability in handling client traffic even in headless mode. To enable this self election, configure the following command under the Service Engine group on the Controller CLI:

```
[admin:avi-cntrlr]: > configure serviceenginegroup Default-Group
[admin:avi-cntrlr]: serviceenginegroup> self_se_election
Overwriting the previously entered value for self_se_election
```

Notes:

1. This feature is supported only for active/active HA deployments in VMware no access and bare metal clouds. 2. The configuration of this option is not yet supported via the Avi UI.

Security

- **HSM Group ?** Hardware security modules may be configured by navigating to Templates > Security > HSM Groups . An HSM is an external security appliance used for secure storage of SSL certificates and keys. An HSM group dictates how Service Engines can reach and authenticate with the HSM. See [Physical Security for SSL Keys.](#)

Log Throttle Settings

- Significant Log Throttle ? This setting limits the number of significant logs generated per second per core on SEs within the group. Default is 100. Set it to zero to disable throttling.
- UDF Log Throttle ? This setting limits the number of UDF (user-defined) logs generated per second per core on SEs within the group. UDF logs are generated due to the configured client log filters or the rules with logging enabled. Default is 100. Set it to zero to disable throttling.
- Non-significant Log Throttle ? This setting limits the number of non-significant logs generated per second per core on SEs within the group. Default is 100. Set it to zero to disable throttling.

Additional Information

[How to Perform an Include/Exclude Operation on a Cluster and a Host at the same time for an SE group](#)