Streaming Gateway Appliance

Streaming Gateway Appliance (SGA) Installation, Upgrade, Management

- Streaming Gateway Appliance
- SGA 4
- SGA 4 Installation
- SGA 4 Upgrade
- SGA 4 Management

Streaming Gateway Appliance

Organizations who have users on the Internet must determine how their users will access their Frame workload VMs in a private network. For these use cases, organizations can provide their users with corporate VPN access or deploy the Frame Streaming Gateway Appliance (SGA), a secure reverse proxy that supports the Frame Remoting Protocol (FRP). SGA enables organizations to grant their users secure access to their virtualized applications and/or desktops without the use of a VPN.

Considerations

Frame provides two options for deploying one or more SGAs. Administrators should review the following considerations to determine which deployment approach fits their requirements.

	Manual SGA Deployment	**Auto SGA Deployment**
Infrastructure	Required for AHV-based accounts. Supported for public cloud accounts.	Supported for public cloud accounts.
Networking	Requires customer-managed networking.	Requires Frame-managed networking.

1. **Auto SGA Deployment:** Frame will provision all of the required network resources (e.g., SGA VPC, security groups/firewall rules, SGA VM(s), VPC/VNET peer, and SGA VMs. SGA VMs will have public IP addresses. Frame will also provision, for SGA 3.X only, a load balancer if more than one SGA VM is required).
2. **Manual SGA Deployment:** Customers manually deploy and register their SGAs and configure their networking/firewall rules. They also provision and configure, for SGA 3.X only, a load balancer if more than one SGA VM is required.

Manual SGA deployments are required when customers have specifc networking requirements (e.g., inbound firewall, WAF, and/or load balancer requirements, prohibition on workload VMs having public IP addresses, outbound NAT or zero-trust Internet access requirements, etc.) that Auto SGA deployment cannot satisfy. In these scenarios, the customer ensures that all SGA and Frame VM network prerequisites are satisfied in order for users to be able to access their workload VMs via a manually deployed SGA Cluster.

IMPORTANT: Upgrading from one SGA version to the next version requires termination and recreation of the SGA VMs. Scheduled downtime may be required.

SGA Versions

SGA Version	Considerations
SGA 3 (out-of-support)	• Supports FRP7 and FRP8. • Requires DNS A record with a wildcard SGA domain name, TLS/SSL public key certificate with the wildcard SGA domain name, and load balancer for high-availability deployments. • For FRP8, each SGA VM must be accessible through its own public IP address.
SGA 4	• Supports FRP8 only. • Managed within Frame Console. • Each SGA VM must be accessible through its own public IP address.

SGA VM Sizing

For customers who are manually deploying SGA VMs (customer-managed networking), customers should start with a configuration for each SGA VM:

- 2 vCPUs
- 4 GB RAM

This configuration ensures the VM can support ~1 Gbps bandwidth of Frame Remoting Protocol data. Frame recommends a sizing target of 500 Mbps per 2 vCPUs to allow users to burst their bandwidth consumption.

The total number of concurrent users for the 500 Mbps bandwidth per 2 vCPU budget is dependent on the bandwidth consumed for the Frame sessions. Bandwidth consumption may be estimated based on user workload profiles:

- 1 Mbps per Frame session for office productivity applications, CPU-only VMs, under 30 fps, 2K or less monitors
- 5 Mbps per Frame session for CAD applications, GPU-backed VMs, up to 60 fps, 2K or less monitors

- 10 Mbps or greater per Frame session for video editing/animation/sustained playback, GPU-backed VMs, up to 60 fps, 2K or less monitors

In an office productivity use case, for example, where CPU-only VMs are used with standard 1920 x 1080 displays, the default (2 vCPU, 4 GB RAM) VM configuration could support 500 concurrent users. For 1,000 concurrent users, the same organization would need to leverage at least a 4 vCPU, 8 GB RAM VM. An 8 vCPU, 16 GB RAM VM could support 2,000 concurrent users for this use case.

Customers who are deploying SGA VMs behind a load balancer for high-availability can incrementally add SGA VMs as their Frame bandwidth consumption increases.

Note

Customers manually deploying SGA VMs in public cloud (customer-managed networking) should ensure they select a non-burstable instance type with sufficient network performance. Public cloud providers may constrain CPU utilization and/or restrict network bandwidth with lower cost instance types.

SGA 4

Introduction

Streaming Gateway Appliance (SGA) 4 simplifies the deployment and management of the SGA by eliminating the need for:

- Public key certificates, as HTTPS is no longer used to communicate between Frame Terminal and the SGA.
- Load balancer, as Frame control plane is responsible for load balancing user sessions across an SGA cluster.

SGA 4 supports:

- Self-service features within Frame Console to:
  - View the status of all SGA clusters and SGA nodes (VMs).
  - Manage the lifecycle of both manually and auto-deployed SGA VMs.
  - Attach (and detach) a Frame account to an SGA cluster.
  - Ability to power on and off individual auto-deployed SGA nodes.
  - Add and delete SGA nodes for a given SGA cluster.
- Customers who require outbound traffic to have a different source public IP address than the inbound public IP address of each SGA VM.
- Customers who require their SGA VMs to have two virtual network interfaces (a virtual network interface for traffic between users on the Internet and the SGA VM and a private virtual network interface for traffic between the SGA VM and the workload VMs).

SGA 4 can only be used with Frame Remoting Protocol (FRP) 8.

Limitations

- SGA 4 image for ESXi will be supported in a future release. SGA 4 is now Generally Available for all other infrastructures.

SGA 4 Clusters and Nodes

SGA 4 introduces two new concepts for customers: SGA Cluster and SGA Node. An SGA Cluster is composed of one or more SGA Nodes, where each node is an SGA 4 VM. Each SGA Cluster is deployed in a specific public cloud region to support Frame Accounts in that region or deployed on-premises to support one or more AHV VLANs. Customers may deploy more than one SGA Cluster. A Frame Account may only be associated to one SGA Cluster.

Once an SGA cluster with one or more nodes is created with at least one node powered on, customers can then in Frame Console:

1. Create a new Frame Account, specifying an existing SGA Cluster.
2. Attach a previously created Frame Account to an existing SGA Cluster (to be supported in a future release).
3. Detach a Frame Account from its SGA Cluster to ensure users can only access the Frame workload VMs in a private networking deployment model (to be supported in a future release).

Network Requirements

When deploying an SGA VM, the customer's network must: (1) allow Internet traffic to reach the SGA VM and (2) from the SGA VM to the network containing the Frame-managed workloads (e.g., Sandbox, test/production pools, Utility Servers). As a best practice, we recommend the SGA VM or VMs (if high availability is required) be deployed in a DMZ (e.g., VPC, VNET, or VLAN) network, separate from the workload VM network.

Customers who have previously configured their network for SGA 3.X will need to allow SGA 4 VMs to initiate outbound HTTPS/Secure WebSocket (tcp/443) connections to cch.console.nutanix.com for communication with Frame control plane.

Customers who are starting with a new network will need to configure their network to satisfy the Frame private networking with SGA 4 network requirements.

Consult Public Cloud with Private Networking and SGA or Nutanix AHV with Private Networking and SGA to ensure that network requirements are satisfied before continuing to SGA 4 installation and configuration.

SGA 4 VM provides Frame Platform its public IP address based on the following:

1. For automated deployments of SGA 4, the public IP address returned from the cloud provider's Instance Metadata Service (IMDS) endpoint.
2. For manual deployments of SGA 4, the public IP address specified by the administrator before the SGA Node is registered to the Frame control plane using the Registration

Code.

> **NOTE**
> 1. While each SGA VM must have an associated public IP address, the public IP address does not have to be attached to virtual network interface of the SGA VM itself. Instead, the customer administrator can manually deploy an SGA VM with only a private IP addresses and then configure a NAT rule either on their firewall, web application firewall, or load balancer that maps the inbound public IP address of the SGA VM to the corresponding private IP address on the SGA VM.
>
> 2. SGA 4 no longer requires a corresponding DNS A record for its public IP address; however, customers can create DNS records for their SGA 4 public IP addresses, if desired.
>
> 3. SGA 4 does not support IPv6 addresses.

High Availability

With SGA 4, Frame control plane will handle load balancing user session requests across the available SGA nodes in the SGA cluster. A load balancer is no longer needed to perform the load balancing function.

High Availability SGA 4 Architecture (FRP8)

High Availability SGA 4 Architecture (FRP8)

Typical FRP8 Workflow

Frame users log in to the Frame Platform and are directed to their Launchpad. When a user clicks the desktop or an application icon in their Launchpad, Frame Platform provides the user's browser with the public IP address of the SGA VM associated with the Frame account.

The user's browser or Frame App begins communicating directly with the specific SGA VM using the provided public IP address using (or). The SGA VM validates the session start request and then forwards the session start request to the user's assigned Frame workload VM using . The Frame Agent on the workload VM validates the session start request and begins the Frame session video/audio stream. FRP8 traffic flows back from the Frame Agent on the workload VM through the SGA VM to the user's browser or Frame App.

Internal Access to SGA-enabled Workloads

SGA 4 also supports the scenario where end users within the private network access the workload VMs of an SGA-enabled Frame account while users on the Internet are accessing workload VMs trough the SGA.

During the WebRTC Interactive Connectivity Establishment (ICE) candidate exchange between user and workload VM, FRP8 will test all ICE candidate pairs and determine the best ICE candidate pair to use. If WebRTC verifies that the user and workload VM can communicate over an internal network path, then the FRP8 stream will use that internal network path.

For internal access by users to the workload VMs of an SGA-enabled Frame account, ensure that the users within their private network can route their traffic to the workload VMs in the private network following the private networking requirements for private networking (public cloud) or private networking (AHV) between the end user and workload VMs.

Multi-Frame Account Support

An SGA 4 cluster can be configured for one or more Frame accounts. If there are Frame accounts in different regions or data centers, we recommend you deploy SGA 4 clusters in each of those different regions or data centers to minimize unnecessary network latency.

Security

SGA 4 appliances use Ubuntu 22.04.3 LTS, hardened using CIS Level 1 Server profile (https://ubuntu.com/security/certifications/docs/usg/cis/compliance). SGA administrators can only access the SGA VM command line only through the infrastructure console. SSH is disabled.

The following ports are bound on the SGA VMs:

- 3478 – (udp/tcp) for FRP8
- 4369 – restricted to localhost requests only by SGA component
- 53 – (udp/tcp) restricted to localhost requests only for Ubuntu systemd-resolve service (DNS)

When a user connects to an SGA 4 node, SGA validates the user session request by confirming the validity of the request with the Frame control plane, before connecting the user with the assigned workload VM.

All communication between the SGA 4 VM and the Frame control plane is conducted using a Secure WebSocket (WSS) connection. The WSS connection is initiated by the SGA 4 VM using HTTPS. During the registration process, the SGA 4 VM will authenticate itself to the Frame control plane using a registration code generated by the control plane (and manually entered by the customer administrator for manually deployed SGA VMs) and provide the SGA 4 VM-specific metadata (UUID, SGA public-private key pair, SGA VM public IP address). Once the Secure WebSocket connection is established, the Frame control plane can communicate with the SGA VM to broker new user sessions, facilitate FRP8 WebRTC negotiation, and monitor the availability of the SGA VM.

The public/private key pair is used by the SGA VM to authenticate itself to Frame control plane each time the SGA VM needs to establish a Secure WebSocket connection to the Frame control plane. The private key is used to sign the initial HTTPS GET request by the SGA VM and the digital signature is sent as one of the HTTPS headers, including the timestamp, UUID, and nonce. The control plane validates the digital signature using the SGA VM public key before agreeing to switch to a Secure WebSocket for bidirectional communication.

SGA 4 Installation

To deploy an SGA 4 Cluster, first decide if you will have Frame automatically deploy the SGA Cluster and Nodes (public cloud only) or you will manually deploy the SGA Nodes (public cloud or Nutanix AHV) of an SGA Cluster yourself.

1. For **automatic deployment**, Frame will handle provisioning of all required public cloud resources in the public cloud region you designate.
   - Frame will provision the VNET/VPC, subnets, security groups, gateways, and requested number of SGA VMs.
   - When a Frame account is created using Frame-managed networking, Frame will peer the SGA VNET/VPC to the workload VM VNET/VPC. For IBM Cloud VPC, Frame will provision a Transit Gateway to connect the two VPCs.
2. For **manual deployment**, you will create the SGA Cluster in Frame Console and then obtain an SGA Node Registration Code for each SGA Node you wish to create for the cluster. You will then enter the SGA Node Registration Code when you provision the SGA VM in your infrastructure console. This registration process enables Frame Platform to know the association between the new SGA Node and the SGA Cluster in Frame.
   - The customer must provision the required network resources (e.g., VNET/VPC, subnets, security groups, gateways) to hold the SGA VMs and then provision the desired number of SGA VMs.
   - When a Frame account is created using customer-managed networking, the customer must peer the network containing the SGA VMs with the customer-managed network containing the workload VMs.

Once an SGA cluster has at least one available SGA node, you will then be able to create a new Frame Account referencing that SGA cluster and/or attach an existing Frame Account to that SGA cluster.

> For public cloud, make sure that you create the SGA Cluster in the same region as the Frame Accounts you wish to attach to the SGA Cluster. For Nutanix AHV, make sure the SGA Cluster is in the same data center as the Frame Accounts you wish to attach to the SGA Cluster. If you do not, then users may experience unacceptable latency, limited bandwidth, and high packet loss, resulting in poor end user experience.

Automatic Deployment

For Automatic Deploy, follow the procedure described under Create Cluster, Automatic Deployment.

Manual Deployment

For Manual Deployment, follow the procedure:

1. Create Cluster, Manual Deployment
2. Add Node, Manual Deployment
3. Add additional nodes following Step 2 as desired.

SGA 4 Upgrade

Administrators need to schedule a maintenance window to upgrade their SGA VM(s) to ensure users know not to access the SGA-enabled workload VMs while the SGA upgrade is in progress.

Administrators can use the Maintenance Mode feature to alert users that the account is undergoing maintenance.

> **Caution**
> The time to perform an SGA upgrade will depend on the infrastructure your account is using, infrastructure traffic, and the number of SGA Nodes to be deployed.

Automatically Deployed SGA Nodes

To upgrade your SGA 4 VMs, add new SGA nodes. Once the new SGA nodes are **Available** under **Streaming Gateways** page for the SGA Cluster, power off the old SGA 4 nodes to test the new SGA 4 nodes. If those new SGA 4 nodes work, then delete the old SGA 4 nodes.

Manually Deployed SGA Nodes

Add new SGA node(s) for the existing SGA Cluster in **Streaming Gateways** page to obtain the Registration Code(s). Then manually provision new SGA node(s) using those Registration Code(s).

Once the new SGA node(s) are availabe, power off the old SGA 4 nodes in your cloud infrastructure console to test the new SGA 4 nodes. If those new SGA 4 nodes work, then delete the old SGA 4 nodes from your cloud infrastructure console.

SGA 4 Management

Management of SGA 4 Nodes and Clusters is on the **Streaming Gateway** page at either the **Customer** or **Organization** entity level. The SGA management functionality will depend on whether you have deployed the SGA Cluster using Frame (Automatic Deployment) or manually (Manual Deployment).

Automatic Deployment

With automatic deployment of an SGA Cluster, Frame is responsible for the lifecycle of all network resources and the SGA VMs. The Frame Accounts must have been created using Frame-managed networking in order for administrators to use Automatic Deployment of SGA. If a Frame account was created using customer-managed networking, then the administrator must manually deploy the SGA cluster and nodes following the instructions under Manual Deployment .

> If you are creating an SGA 4 cluster with the expectation of upgrading from existing SGA 3.x Frame accounts, please ensure you use a non-overlapping CIDR for the SGA 4 cluster. To accomplish this, enable the "Use custom CIDR range" slider in the Create Streaming Gateway Cluster configuration form.

Create Cluster

1. To create a new SGA cluster, go to the Frame Console and at the Frame Customer or Organization entity level, click on **Streaming Gateways** on the lefthand menu.
2. Click on **Create New Cluster** in the upper right corner.

image.png

3. Select “Automatic” (Frame creates all resources) and then click the **Continue** button.

4. Complete the Create Streaming Gateway configuration form.

image.png

- **Name:** Name of the SGA cluster. The name of each SGA node will be the SGA cluster name appended with a unique ID.
- **Cloud Provider:** Select the cloud provider you wish to use for this SGA cluster. -
- **Cloud Account:** Select the Cloud Account where the public cloud resources for this cluster will be provisioned. -
- **Region:** Select the cloud region where the SGA cluster will reside.
- **Number of VMs:** Specify the number of SGA nodes (VMs) to be provisioned.
- **Custom CIDR:** Specify the CIDR range where the SGA nodes will be provisioned (default 172.16.0.0/24).

5. Once the required field values have been specified, click on the **Create** button to create the SGA Cluster and the SGA Nodes. You can view the status of the SGA Cluster on the Streaming Gateways page.

After your SGA cluster and SGA nodes have been created, you can then reference the SGA Cluster when creating your Frame Account so that your newly created Frame account uses the SGA Cluster.

Delete Cluster

A SGA Cluster can be deleted only if there are no Frame Accounts attached to the cluster.

1. To delete an SGA cluster, go to the Frame Console and at the Frame Customer or Organization entity level where the SGA Cluster is defined, click on **Streaming Gateways** on the lefthand menu.
2. Click on the kebab menu to the right of the SGA Cluster and select **Delete**.

image.png

3. You will be asked to confirm that you wish to delete the SGA cluster. Click **Cancel** or **Delete**.

image.png

For SGA 4 clusters that were automatically deployed, Frame Console will terminate the SGA Nodes and the related SGA network resources (subnets, VPC/VNET) in the infrastructure and then delete the SGA Cluster.

Add Node

Customer administrators can add another SGA 4 Node to their SGA Cluster at any time.

1. For automatically deployed SGA Clusters, navigate to the **Streaming Gateways** page and locate the SGA Cluster you wish to add a new SGA Node.

   image.png

2. Click on **+ Add a Node**. Frame will provision a new SGA VM and wait for the VM to register.

   image.png

3. The Status of the SGA Node will change from `Pending registration` to `Available` once the SGA Node registers.

> **SGA Instance types**
> Frame Platform will provision SGA VM(s) on the following instance/machine types. These VMs will run 24x7 since users need to be able to access the workload VMs at any time. Administrators can manually power off and power on SGA VMs that are auto deployed.
>
> - **AWS:** c5.xlarge, 30 GB disk
> - **Azure:** D4 v3, 30 GB disk
> - **GCP:** e2-standard-4, 50 GB disk
> - **IBM:** cx3d-4x10, 130 GB disk

Delete Node

Customer administrators can delete an existing SGA 4 Node from their SGA Cluster at any time, except when:

- The SGA 4 node is powered on.
- There is only one SGA 4 node left and there are one or more Frame accounts attached to the SGA Cluster.

1. For automatically deployed SGA Clusters, navigate to the **Streaming Gateways** page and locate the SGA Cluster you wish to delete one of the existing SGA Nodes. Note that the SGA Node to be deleted must be powered off with status `Unavailable`.

   image.png

2. Click on the kebab menu for the powered off SGA node and select **Delete**.

image.png

3. Confirm that you wish to delete the SGA Node by clicking on the **Delete node** button.

image.png

4. Once the SGA Node is deleted, the SGA Node will be removed from the list of nodes for the SGA Cluster.

image.png

Power On Node

With automatically deployed SGA 4 Clusters, customer administrators can power on an existing automatically deployed SGA 4 Node, if the VM is powered off.

1. Navigate to the **Streaming Gateways** page and locate the SGA Cluster containing one or more SGA Nodes you wish to power on.

image.png

2. Click on the kebab menu and select **Start**.

image.png

3. You will be asked to confirm that you wish to power on the SGA Node.

image.png

Power Off Node

With automatically deployed SGA 4 Clusters, customer administrators can power off an existing automatically deployed SGA 4 Node, if the VM is powered on.

1. Navigate to the **Streaming Gateways** page and locate the SGA Cluster containing one or more SGA Nodes you wish to power off.

   image.png

2. Click on the kebab menu and select **Stop**.

   image.png

3. You will be asked to confirm that you wish to power off the SGA Node.

   image.png

Reboot Node

With automatically deployed SGA 4 Clusters, customer administrators can reboot an existing automatically deployed SGA 4 Node.

1. Navigate to the **Streaming Gateways** page and locate the SGA Cluster containing one or more SGA Nodes you wish to power off.

   image.png

2. Click on the kebab menu and select **Reboot**.

   image.png

3. You will be asked to confirm that you wish to reboot the SGA Node.

   image.png

Manual Deployment

With manual deployment of an SGA Cluster, the customer is responsible for the lifecycle of all network resources and the SGA VMs. The Frame Accounts must have been created using customer-managed networking in order for administrators to use Manual Deployment of SGA. If a Frame account was created using Frame-managed networking, then the administrator must follow the instructions under Automatic Deployment of an SGA Cluster.

Create Cluster

1. To create a new SGA cluster, go to the Frame Console and at the Frame Customer or Organization entity level, click on **Streaming Gateways** on the lefthand menu.
2. Click on **Create New Cluster** in the upper right corner.

3. Select “Manual” and then click the **Continue** button. 4. Complete the Create Streaming Gateway configuration form.

image.png

- **Name:** Name of the SGA cluster. The name of each SGA node will be the SGA cluster name appended with a unique ID.
- **Cloud Provider:** Select the cloud provider you wish to use for this SGA cluster.
- **Cloud Account:** Select the Cloud Account where the public cloud resources for this cluster will be provisioned.

5. Frame Console will display the newly created SGA Cluster and one SGA Node entry. Notice that Frame Console provides the Activation Code for this first SGA Node. You will need this Activation Code to register the SGA Node you manually provision.

image.png

Once the SGA Cluster has at least one registered SGA Node, you can then reference the SGA cluster when creating new Frame accounts.

> The Activation Code must be used within 30 minutes of creation. If the Activation Code expires, you may click on **Generate new code** on the SGA Node line to obtain a new Activation Code.

Delete Cluster

A SGA Cluster can be deleted only if there are no Frame Accounts attached to the cluster.

1. To delete an SGA cluster, go to the Frame Console and at the Frame Customer or Organization entity level where the SGA Cluster is defined, click on **Streaming Gateways** on the lefthand menu.
2. Click on the kebab menu to the right of the SGA Cluster and select **Delete**.

image.png

3. You will be asked to confirm that you wish to delete the SGA cluster. Click **Cancel** or **Delete**.

If the SGA 4 cluster was manually deployed, Frame Console will delete the SGA Cluster in Frame Console. However, the customer is responsible for terminating the SGA Nodes and any related SGA network resources in their infrastructure.

Add Node

For customers who are manually deploying an SGA Cluster, you must manually provision and configure the SGA Nodes from within your infrastructure console.

Prerequisites

SGA 4 prerequisites are as follows:

- Download the Frame SGA disk image from the Downloads Page for the hypervisor/infrastructure on which you wish to deploy the SGA.
- Determine which data center or public cloud region where the SGA Nodes for the SGA Cluster will be provisioned. To minimize network latency for the best user experience, the SGA Nodes for an SGA Cluster and the associated Frame accounts using that SGA Cluster should be in the same data center or public cloud region.
- Configure the firewall(s) and networking to support the required FRP8 protocols/ports from the Internet to the SGA Cluster and from the SGA Cluster to the workload network (e.g., VLAN or VNET/VPC and subnet) as well as from the workload network back to the Internet via the SGA Cluster.
- Assign a static private IP address to each SGA VM.
- Assign a static public IP address to each SGA VM. The public IP address can be configured in a firewall or load balancer with network address translation (NAT) to the SGA VM private IP address.

Step 1: Provision the SGA Node

1. To manually create an SGA Node, go to the **Streaming Gateways** page and find the Manually Deployed SGA Cluster that will have this new SGA Node. You may need to click on **>>**