Manage load balancing

StorageGRID 11.7

NetApp
January 09, 2024

Table of Contents

Manage load balancing

Considerations for load balancing

You can use load balancing to handle ingest and retrieval workloads from S3 and Swift clients.

What is load balancing?

When a client application saves or retrieves data from a StorageGRID system, StorageGRID uses a load balancer to manage the ingest and retrieval workload. Load balancing maximizes speed and connection capacity by distributing the workload across multiple Storage Nodes.

The StorageGRID Load Balancer service is installed on all Admin Nodes and all Gateway Nodes and provides Layer 7 load balancing. It performs Transport Layer Security (TLS) termination of client requests, inspects the requests, and establishes new secure connections to the Storage Nodes.

The Load Balancer service on each node operates independently when forwarding client traffic to the Storage Nodes. Through a weighting process, the Load Balancer service routes more requests to Storage Nodes with higher CPU availability.

> Although the StorageGRID Load Balancer service is the recommended load balancing mechanism, you might want to integrate a third-party load balancer instead. For information, contact your NetApp account representative or refer to TR-4626: StorageGRID third-party and global load balancers.

How many load balancing nodes do I need?

As a general best practice, each site in your StorageGRID system should include two or more nodes with the Load Balancer service. For example, a site might include two Gateway Nodes or both an Admin Node and a Gateway Node. Make sure that there is adequate networking, hardware, or virtualization infrastructure for each load-balancing node, whether you are using SG100 or SG1000 services appliances, bare metal nodes, or virtual machine (VM) based nodes.

What is a load balancer endpoint?

A load balancer endpoint defines the port and the network protocol (HTTPS or HTTP) that incoming and outgoing client application requests will use to access those nodes that contain the Load Balancer service. The endpoint also defines the client type (S3 or Swift), the binding mode, and optionally a list of allowed or blocked tenants.

To create a load balancer endpoint, either select **CONFIGURATION** > **Network** > **Load balancer endpoints** or complete the FabricPool and S3 setup wizard. For instructions:

- Configure load balancer endpoints
- Use the S3 setup wizard
- Use the FabricPool setup wizard

Considerations for the port

The port for a load balancer endpoint defaults to 10433 for the first endpoint you create, but you can specify any unused external port between 1 and 65535. If you use port 80 or 443, the endpoint will use the Load Balancer service on Gateway Nodes only. These ports are reserved on Admin Nodes. If you use the same port for more than one endpoint, you must specify a different binding mode for each endpoint.

Ports used by other grid services aren't permitted. See the Network port reference.

Considerations for the network protocol

In most cases, the connections between client applications and StorageGRID should use Transport Layer Security (TLS) encryption. Connecting to StorageGRID without TLS encryption is supported but not recommended, especially in production environments. When you select the network protocol for the StorageGRID load balancer endpoint, you should select **HTTPS**.

Considerations for load balancer endpoint certificates

If you select **HTTPS** as the network protocol for the load balancer endpoint, you must provide a security certificate. You can use any of these three options when you create the load balancer endpoint:

- **Upload a signed certificate (recommended)**. This certificate can be signed by either a publicly trusted or a private certificate authority (CA). Using a publicly trusted CA server certificate to secure the connection is the best practice. In contrast to generated certificates, certificates signed by a CA can be rotated nondisruptively, which can help avoid expiration issues.

  You must obtain the following files before you create the load balancer endpoint:

  - The custom server certificate file.
  - The custom server certificate private key file.
  - Optionally, a CA bundle of the certificates from each intermediate issuing certificate authority.
- **Generate a self-signed certificate**.
- **Use the global StorageGRID S3 and Swift certificate**. You must upload or generate a custom version of this certificate before you can select it for the load balancer endpoint. See Configure S3 and Swift API certificates.

What values do I need?

To create the certificate, you must know all of the domain names and IP addresses that S3 or Swift client applications will use to access the endpoint.

The **Subject DN** (Distinguished Name) entry for the certificate must include the fully qualified domain name that the client application will use for StorageGRID. For example:

```
Subject DN:
/C=Country/ST=State/O=Company,Inc./CN=s3.storagegrid.example.com
```

As required, the certificate can use wildcards to represent the fully qualified domain names of all Admin Nodes and Gateway Nodes running the Load Balancer service. For example, `*.storagegrid.example.com` uses the * wildcard to represent `adm1.storagegrid.example.com` and `gn1.storagegrid.example.com`.

If you plan to use S3 virtual hosted-style requests, the certificate must also include an **Alternative Name** entry for each S3 endpoint domain name you have configured, including any wildcard names. For example:

```
Alternative Name: DNS:*.s3.storagegrid.example.com
```

If you use wildcards for domain names, review the Hardening guidelines for server certificates.

You must also define a DNS entry for each name in the security certificate.

How do I manage expiring certificates?

If the certificate used to secure the connection between the S3 application and StorageGRID expires, the application might temporarily lose access to StorageGRID.

To avoid certificate expiration issues, follow these best practices:

- Carefully monitor any alerts that warn of approaching certificate expiration dates, such as the **Expiration of load balancer endpoint certificate** and **Expiration of global server certificate for S3 and Swift API** alerts.
- Always keep the StorageGRID and S3 application's versions of the certificate in sync. If you replace or renew the certificate used for a load balancer endpoint, you must replace or renew the equivalent certificate used by the S3 application.
- Use a publicly signed CA certificate. If you use a certificate signed by a CA, you can replace soon-to-expire certificates nondisruptively.
- If you have generated a self-signed StorageGRID certificate and that certificate is about to expire, you must manually replace the certificate in both StorageGRID and in the S3 application before the existing certificate expires.

Considerations for the binding mode

The binding mode lets you control which IP addresses can be used to access a load balancer endpoint. If an endpoint uses a binding mode, client applications can only access the endpoint if they use an allowed IP address or its corresponding fully qualified domain name (FQDN). Client applications using any other IP address or FQDN can't access the endpoint.

You can specify any of the following binding modes:

- **Global** (default): Client applications can access the endpoint using the IP address of any Gateway Node or Admin Node, the virtual IP (VIP) address of any HA group on any network, or a corresponding FQDN. Use this setting unless you need to restrict the accessibility of an endpoint.
- **Virtual IPs of HA groups**. Client applications must use a virtual IP address (or corresponding FQDN) of an HA group.
- **Node interfaces**. Clients must use the IP addresses (or corresponding FQDNs) of selected node interfaces.
- **Node type**. Based on the type of node you select, clients must use either the IP address (or corresponding FQDN) of any Admin Node or the IP address (or corresponding FQDN) of any Gateway Node.

Considerations for tenant access

Tenant access is an optional security feature that lets you control which StorageGRID tenant accounts can use a load balancer endpoint to access their buckets. You can allow all tenants to access an endpoint (default), or you can specify a list of the allowed or blocked tenants for each endpoint.

You can use this feature to provide better security isolation between tenants and their endpoints. For example, you might use this feature to ensure that the top-secret or highly classified materials owned by one tenant remain completely inaccessible to other tenants.

> For the purpose of access control, the tenant is determined from the access keys used in the client request, if no access keys are provided as part of the request (such as with anonymous access) the bucket owner is used to determine the tenant.

Tenant access example

To understand how this security feature works, consider the following example:

1. You have created two load balancer endpoints, as follows:
   - **Public** endpoint: Uses port 10443 and allows access to all tenants.
   - **Top secret** endpoint: Uses port 10444 and allows access to the **Top secret** tenant only. All other tenants are blocked from accessing this endpoint.
2. The `top-secret.pdf` is in a bucket owned by the **Top secret** tenant.

To access the `top-secret.pdf`, a user in the **Top secret** tenant can issue a GET request to `https://w.x.y.z:10444/top-secret.pdf`. Because this tenant is allowed to use the 10444 endpoint, the user can access the object. However, if a user belonging to any other tenant issues the same request to the same URL, they receive an immediate Access Denied message. Access is denied even if the credentials and signature are valid.

CPU availability

The Load Balancer service on each Admin Node and Gateway Node operates independently when forwarding S3 or Swift traffic to the Storage Nodes. Through a weighting process, the Load Balancer service routes more requests to Storage Nodes with higher CPU availability. Node CPU load information is updated every few minutes, but weighting might be updated more frequently. All Storage Nodes are assigned a minimal base weight value, even if a node reports 100% utilization or fails to report its utilization.

In some cases, information about CPU availability is limited to the site where the Load Balancer service is located.

Configure load balancer endpoints

Load balancer endpoints determine the ports and network protocols S3 and Swift clients can use when connecting to the StorageGRID load balancer on Gateway and Admin Nodes.

> Support for Swift client applications has been deprecated and will be removed in a future release.

Before you begin

- You are signed in to the Grid Manager using a supported web browser.
- You have the Root access permission.
- You have reviewed the considerations for load balancing.
- If you previously remapped a port you intend to use for the load balancer endpoint, you have removed the port remap.
- You have created any high availability (HA) groups you plan to use. HA groups are recommended, but not required. See Manage high availability groups.
- If the load balancer endpoint will be used by S3 tenants for S3 Select, it must not use the IP addresses or FQDNs of any bare-metal nodes. Only SG100 or SG1000 appliances and VMware-based software nodes are allowed for the load balancer endpoints used for S3 Select.
- You have configured any VLAN interfaces you plan to use. See Configure VLAN interfaces.
- If you are creating an HTTPS endpoint (recommended), you have the information for the server certificate.

  Changes to an endpoint certificate can take up to 15 minutes to be applied to all nodes.

  - To upload a certificate, you need the server certificate, the certificate private key, and optionally, a CA bundle.
  - To generate a certificate, you need all of the domain names and IP addresses that S3 or Swift clients will use to access the endpoint. You must also know the subject (Distinguished Name).
  - If you want to use the StorageGRID S3 and Swift API certificate (which can also be used for connections directly to Storage Nodes), you have already replaced the default certificate with a custom certificate signed by an external certificate authority. See Configure S3 and Swift API certificates.

Create a load balancer endpoint

Each load balancer endpoint specifies a port, a client type (S3 or Swift), and a network protocol (HTTP or HTTPS).

Access the wizard

Steps

1. Select **CONFIGURATION** > **Network** > **Load balancer endpoints**.
2. Select **Create**.

Enter endpoint details

Steps

1. Enter details for the endpoint.

Field	Description
Name	A descriptive name for the endpoint, which will appear in the table on the Load balancer endpoints page.

Field	Description
Port	The StorageGRID port you want to use for load balancing. This field defaults to 10433 for the first endpoint you create, but you can enter any unused external port between 1 and 65535. If you enter **80** or **443**, the endpoint is configured only on Gateway Nodes. These ports are reserved on Admin Nodes.
Client type	The type of client application that will use this endpoint, either **S3** or **Swift**.
Network protocol	The network protocol that clients will use when connecting to this endpoint. • Select **HTTPS** for secure, TLS encrypted communication (recommended). You must attach a security certificate before you can save the endpoint. • Select **HTTP** for less secure, unencrypted communication. Use HTTP only for a non-production grid.

2. Select **Continue**.

Select a binding mode

Steps

1. Select a binding mode for the endpoint to control how the endpoint is accessed—using any IP address or using specific IP addresses and network interfaces.

Option	Description
Global (default)	Clients can access the endpoint using the IP address of any Gateway Node or Admin Node, the virtual IP (VIP) address of any HA group on any network, or a corresponding FQDN. Use the **Global** setting (default) unless you need to restrict the accessibility of this endpoint.
Virtual IPs of HA groups	Clients must use a virtual IP address (or corresponding FQDN) of an HA group to access this endpoint. Endpoints with this binding mode can all use the same port number, as long as the HA groups you select for the endpoints don’t overlap.
Node interfaces	Clients must use the IP addresses (or corresponding FQDNs) of selected node interfaces to access this endpoint.
Node type	Based on the type of node you select, clients must use either the IP address (or corresponding FQDN) of any Admin Node or the IP address (or corresponding FQDN) of any Gateway Node to access this endpoint.

If more than one endpoint uses the same port, StorageGRID uses this priority order to decide which endpoint to use: **Virtual IPs of HA groups** > **Node interfaces** > **Node type** > **Global**.

2. If you selected **Virtual IPs of HA groups**, select one or more HA groups.
3. If you selected **Node interfaces**, select one or more node interfaces for each Admin Node or Gateway Node that you want to associate with this endpoint.
4. If you selected **Node type**, select either Admin Nodes, which includes both the primary Admin Node and any non-primary Admin Nodes, or Gateway Nodes.

Control tenant access

Steps

1. For the **Tenant access** step, select one of the following:

Field	Description
Allow all tenants (default)	All tenant accounts can use this endpoint to access their buckets. You must select this option if you have not yet created any tenant accounts. After you add tenant accounts, you can edit the load balancer endpoint to allow or block specific accounts.
Allow selected tenants	Only the selected tenant accounts can use this endpoint to access their buckets.
Block selected tenants	The selected tenant accounts can't use this endpoint to access their buckets. All other tenants can use this endpoint.

2. If you are creating an **HTTP** endpoint, you don't need to attach a certificate. Select **Create** to add the new load balancer endpoint. Then, go to After you finish. Otherwise, select **Continue** to attach the certificate.

Attach certificate

Steps

1. If you are creating an **HTTPS** endpoint, select the type of security certificate you want to attach to the endpoint.

   The certificate secures the connections between S3 and Swift clients and the Load Balancer service on Admin Node or Gateway Nodes.

   - **Upload certificate**. Select this option if you have custom certificates to upload.
   - **Generate certificate**. Select this option if you have the values needed to generate a custom certificate.
   - **Use StorageGRID S3 and Swift certificate**. Select this option if you want to use the global S3 and Swift API certificate, which can also be used for connections directly to Storage Nodes.

     You can't select this option unless you have replaced the default S3 and Swift API certificate, which is signed by the grid CA, with a custom certificate signed by an external certificate authority. See Configure S3 and Swift API certificates.

2. If you aren’t using the StorageGRID S3 and Swift certificate, upload or generate the certificate.

Upload certificate

a. Select **Upload certificate**.

b. Upload the required server certificate files:

- **Server certificate**: The custom server certificate file in PEM encoding.
- **Certificate private key**: The custom server certificate private key file (`.key`).

> EC private keys must be 224 bits or larger. RSA private keys must be 2048 bits or larger.

- **CA bundle**: A single optional file containing the certificates from each intermediate issuing certificate authority (CA). The file should contain each of the PEM-encoded CA certificate files, concatenated in certificate chain order.

c. Expand **Certificate details** to see the metadata for each certificate you uploaded. If you uploaded an optional CA bundle, each certificate displays on its own tab.

- Select **Download certificate** to save the certificate file or select **Download CA bundle** to save the certificate bundle.

  Specify the certificate file name and download location. Save the file with the extension `.pem`.

  For example: `storagegrid_certificate.pem`

- Select **Copy certificate PEM** or **Copy CA bundle PEM** to copy the certificate contents for pasting elsewhere.

d. Select **Create**.
The load balancer endpoint is created. The custom certificate is used for all subsequent new connections between S3 and Swift clients and the endpoint.

Generate certificate

a. Select **Generate certificate**.

b. Specify the certificate information:

Field	Description
Domain name	One or more fully qualified domain names to include in the certificate. Use an * as a wildcard to represent multiple domain names.
IP	One or more IP addresses to include in the certificate.
Subject (optional)	X.509 subject or distinguished name (DN) of the certificate owner. If no value is entered in this field, the generated certificate uses the first domain name or IP address as the subject common name (CN).
Days valid	Number of days after creation that the certificate expires.

Field	Description
Add key usage extensions	If selected (default and recommended), key usage and extended key usage extensions are added to the generated certificate. These extensions define the purpose of the key contained in the certificate. **Note**: Leave this checkbox selected unless you experience connection problems with older clients when certificates include these extensions.

c. Select **Generate**.

d. Select **Certificate details** to see the metadata for the generated certificate.

- Select **Download certificate** to save the certificate file.

  Specify the certificate file name and download location. Save the file with the extension `.pem`.

  For example: `storagegrid_certificate.pem`

- Select **Copy certificate PEM** to copy the certificate contents for pasting elsewhere.

e. Select **Create**.

The load balancer endpoint is created. The custom certificate is used for all subsequent new connections between S3 and Swift clients and this endpoint.

After you finish

Steps

1. If you use a DNS, ensure that the DNS includes a record to associate the StorageGRID fully qualified domain name (FQDN) to each IP address that clients will use to make connections.

   The IP address you enter in the DNS record depends on whether you are using an HA group of load-balancing nodes:

   - If you have configured an HA group, clients will connect to the virtual IP addresses of that HA group.
   - If you aren’t using an HA group, clients will connect to the StorageGRID Load Balancer service using the IP address of a Gateway Node or Admin Node.

     You must also ensure that the DNS record references all required endpoint domain names, including any wildcard names.

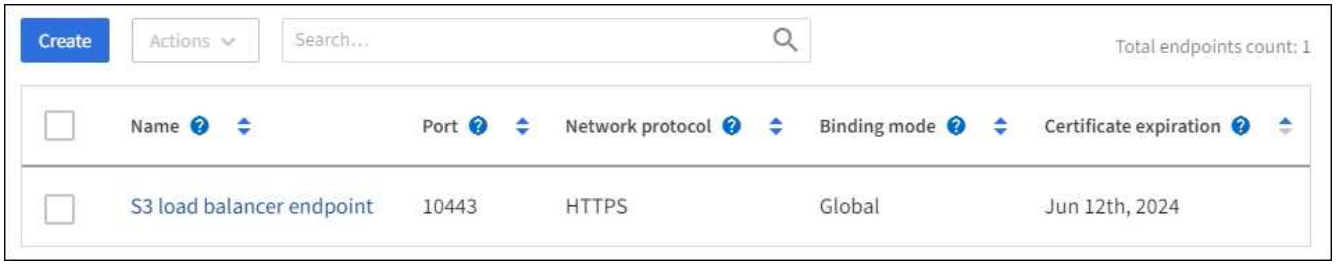
2. Provide S3 and Swift clients with the information needed to connect to the endpoint:
   - Port number
   - Fully qualified domain name or IP address
   - Any required certificate details

View and edit load balancer endpoints

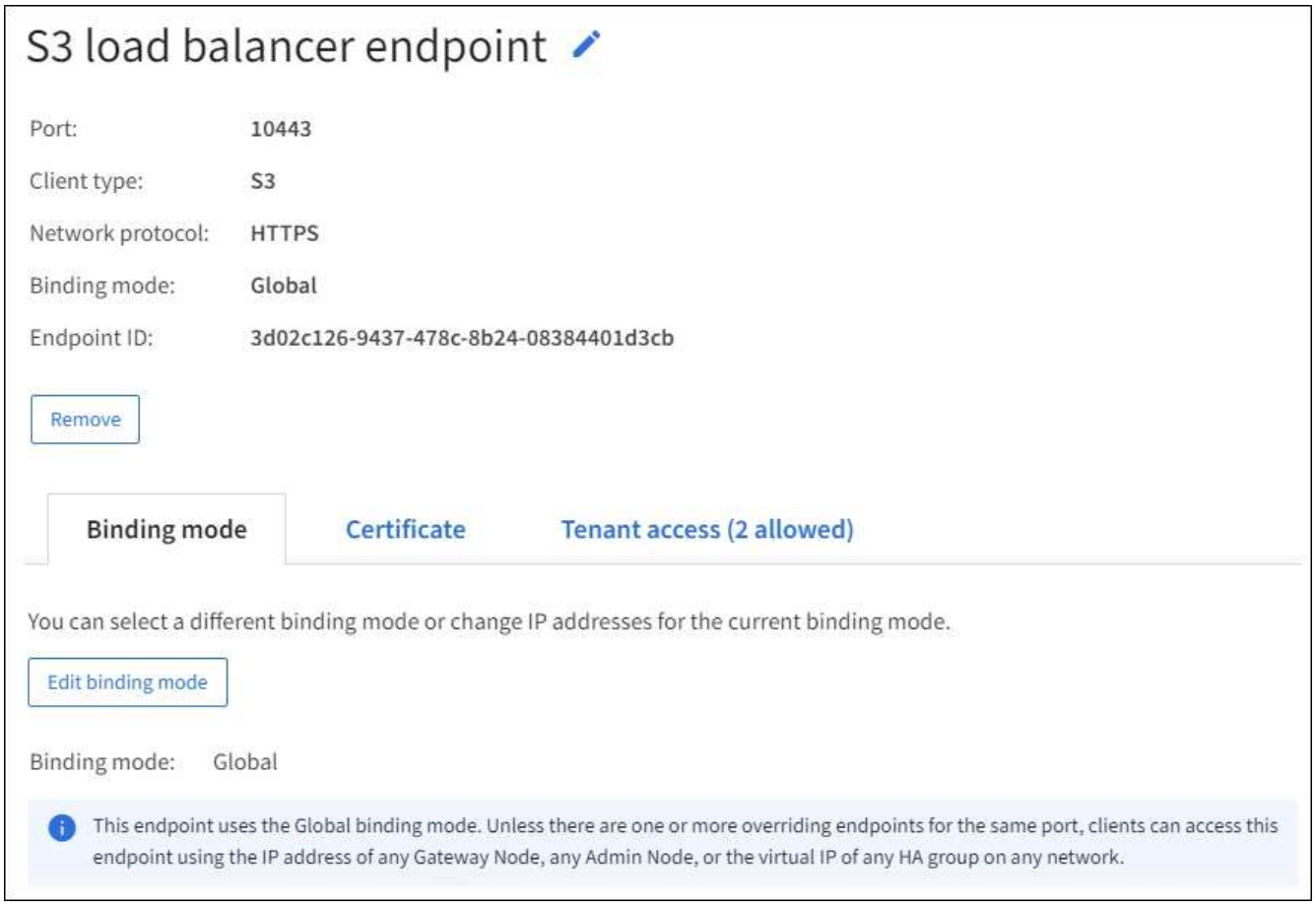
You can view details for existing load balancer endpoints, including the certificate metadata for a secured endpoint. You can also change an endpoint's name or binding mode and update any associated certificates.

You can't change the service type (S3 or Swift), the port, or the protocol (HTTP or HTTPS).

- To view basic information for all load balancer endpoints, review the table on the Load balancer endpoints page.



- To view all details about a specific endpoint, including certificate metadata, select the endpoint's name in the table.



- To edit an endpoint, use the **Actions** menu on the Load balancer endpoints page or the details page for a specific endpoint.

> After editing an endpoint, you might need to wait up to 15 minutes for your changes to be applied to all nodes.

Task	Actions menu	Details page
Edit endpoint name	a. Select the checkbox for the endpoint. b. Select **Actions** > **Edit endpoint name**. c. Enter the new name. d. Select **Save**.	a. Select the endpoint name to display the details. b. Select the edit icon . c. Enter the new name. d. Select **Save**.
Edit endpoint binding mode	a. Select the checkbox for the endpoint. b. Select **Actions** > **Edit endpoint binding mode**. c. Update the binding mode as required. d. Select **Save changes**.	a. Select the endpoint name to display the details. b. Select **Edit binding mode**. c. Update the binding mode as required. d. Select **Save changes**.
Edit endpoint certificate	a. Select the checkbox for the endpoint. b. Select **Actions** > **Edit endpoint certificate**. c. Upload or generate a new custom certificate or begin using the global S3 and Swift certificate, as required. d. Select **Save changes**.	a. Select the endpoint name to display the details. b. Select the **Certificate** tab. c. Select **Edit certificate**. d. Upload or generate a new custom certificate or begin using the global S3 and Swift certificate, as required. e. Select **Save changes**.
Edit tenant access	a. Select the checkbox for the endpoint. b. Select **Actions** > **Edit tenant access**. c. Choose a different access option, select or remove tenants from the list, or do both. d. Select **Save changes**.	a. Select the endpoint name to display the details. b. Select the **Tenant access** tab. c. Select **Edit tenant access**. d. Choose a different access option, select or remove tenants from the list, or do both. e. Select **Save changes**.

Remove load balancer endpoints

You can remove one or more endpoints using the **Actions** menu, or you can remove a single endpoint from the details page.

> To prevent client disruptions, update any affected S3 or Swift client applications before you remove a load balancer endpoint. Update each client to connect using a port assigned to another load balancer endpoint. Be sure to update any required certificate information as well.

- To remove one or more endpoints:
  a. From the Load balancer page, select the checkbox for each endpoint you want to remove.

b. Select **Actions** > **Remove**.

c. Select **OK**.

- To remove one endpoint from the details page:

a. From the Load balancer page. select the endpoint name.

b. Select **Remove** on the details page.

c. Select **OK**.

Copyright information

Trademark information