NetApp

Plan StorageGRID expansion

StorageGRID 11.7

NetApp
March 05, 2024

Table of Contents

Plan StorageGRID expansion

Add storage capacity

Guidelines for adding object capacity

You can expand the object storage capacity of your StorageGRID system by adding storage volumes to existing Storage Nodes or by adding new Storage Nodes to existing sites. You must add storage capacity in a way that meets the requirements of your information lifecycle management (ILM) policy.

Guidelines for adding storage volumes

Before adding storage volumes to existing Storage Nodes, review the following guidelines and limitations:

- You must examine your current ILM rules to determine where and when to add storage volumes to increase the storage available for replicated objects or erasure-coded objects.

- You can't increase the metadata capacity of your system by adding storage volumes because object metadata is stored only on volume 0.

- Each software-based Storage Node can support a maximum of 16 storage volumes. If you need to add capacity beyond that, you must add new Storage Nodes.

- You can add one or two expansion shelves to each SG6060 appliance. Each expansion shelf adds 16 storage volumes. With both expansion shelves installed, the SG6060 can support a total of 48 storage volumes.

- You can't add storage volumes to any other storage appliance.

- You can't increase the size of an existing storage volume.

- You can't add storage volumes to a Storage Node at the same time you are performing a system upgrade, recovery operation, or another expansion.

After you have decided to add storage volumes and have determined which Storage Nodes you must expand to satisfy your ILM policy, follow the instructions for your type of Storage Node:

- To add one or two expansion shelves to an SG6060 storage appliance, go to Add expansion shelf to deployed SG6060.

- For a software-based node, follow the instructions for adding storage volumes to Storage Nodes.

Guidelines for adding Storage Nodes

Before adding Storage Nodes to existing sites, review the following guidelines and limitations:

- You must examine your current ILM rules to determine where and when to add Storage Nodes to increase the storage available for replicated objects or erasure-coded objects.

- You should not add more than 10 Storage Nodes in a single expansion procedure.

- You can add Storage Nodes to more than one site in a single expansion procedure.

- You can add Storage Nodes and other types of nodes in a single expansion procedure.

- Before starting the expansion procedure, you must confirm that all data-repair operations performed as part of a recovery are complete. See Check data repair jobs.

- If you need to remove Storage Nodes before or after performing an expansion, you should not decommission more than 10 Storage Nodes in a single Decommission Node procedure.

Guidelines for ADC service on Storage Nodes

When configuring the expansion, you must choose whether to include the Administrative Domain Controller (ADC) service on each new Storage Node. The ADC service keeps track of the location and availability of grid services.

- The StorageGRID system requires a quorum of ADC services to be available at each site and at all times.

- At least three Storage Nodes at each site must include the ADC service.

- Adding the ADC service to every Storage Node is not recommended. Including too many ADC services can cause slowdowns due to the increased amount of communication between nodes.

- A single grid should not have more than 48 Storage Nodes with the ADC service. This is equivalent to 16 sites with three ADC services at each site.

- In general, when you select the **ADC Service** setting for a new node, you should select **Automatic**. Select **Yes** only if the new node will replace another Storage Node that includes the ADC service. Because you can't decommission a Storage Node if too few ADC services would remain, this ensures that a new ADC service is available before the old service is removed.

- You can't add the ADC service to a node after it is deployed.

Add storage capacity for replicated objects

If the information lifecycle management (ILM) policy for your deployment includes a rule that creates replicated copies of objects, you must consider how much storage to add and where to add the new storage volumes or Storage Nodes.

For guidance on where to add additional storage, examine the ILM rules that create replicated copies. If ILM rules create two or more object copies, plan to add storage in each location where object copies are made. As a simple example, if you have a two-site grid and an ILM rule that creates one object copy at each site, you must add storage to each site to increase the overall object capacity of the grid. For information about object replication, see What is replication?.

For performance reasons, you should attempt to keep storage capacity and compute power balanced across sites. So, for this example, you should add the same number of Storage Nodes to each site or additional storage volumes at each site.

If you have a more complex ILM policy that includes rules that place objects in different locations based on criteria such as bucket name, or rules that change object locations over time, your analysis of where storage is required for the expansion will be similar, but more complex.

Charting how quickly overall storage capacity is being consumed can help you understand how much storage to add in the expansion, and when the additional storage space will be required. You can use the Grid Manager to monitor and chart storage capacity.

When planning the timing of an expansion, remember to consider how long it might take to procure and install additional storage.

Add storage capacity for erasure-coded objects

If your ILM policy includes a rule that makes erasure-coded copies, you must plan where

to add new storage and when to add new storage. The amount of storage you add and the timing of the addition can affect the grid's usable storage capacity.

The first step in planning a storage expansion is to examine the rules in your ILM policy that create erasure-coded objects. Because StorageGRID creates $k+m$ fragments for every erasure-coded object and stores each fragment on a different Storage Node, you must ensure that at least $k+m$ Storage Nodes have space for new erasure-coded data after the expansion. If the erasure coding profile provides site-loss protection, you must add storage to each site. See What are erasure coding schemes? for information about erasure coding profiles.

The number of nodes you need to add also depends on how full the existing nodes are when you perform the expansion.

General recommendation for adding storage capacity for erasure-coded objects

If you want to avoid detailed calculations, you can add two Storage Nodes per site when existing Storage Nodes reach 70% capacity.

This general recommendation provides reasonable results across a wide range of erasure-coding schemes for both single-site grids and for grids where erasure coding provides site-loss protection.

To better understand the factors that led to this recommendation or to develop a more precise plan for your site, see Considerations for rebalancing erasure-coded data. For a custom recommendation optimized for your situation, contact your NetApp Professional Services consultant.

Considerations for rebalancing erasure-coded data

If you are performing an expansion to add Storage Nodes and you use ILM rules to erasure code data, you might need to perform the EC rebalance procedure if you can't add enough Storage Nodes for the erasure-coding scheme you are using.

After reviewing these considerations, perform the expansion, and then go to Rebalance erasure-coded data after adding Storage Nodes to run the procedure.

What is EC rebalancing?

EC rebalancing is a StorageGRID procedure that might be required after a Storage Node expansion. The procedure is run as a command-line script from the primary Admin Node. When you run the EC rebalance procedure, StorageGRID redistributes erasure-coded fragments among the existing and the newly added Storage Nodes at a site.

The EC rebalance procedure:

- Only moves erasure-coded object data. It does not move replicated object data.
- Redistributes the data within a site. It does not move data between sites.
- Redistributes data among all Storage Nodes at a site. It does not redistribute data within storage volumes.
- Does not consider the replicated data usage on each Storage Node when determining where to move erasure-coded data.
- Redistributes erasure-coded data evenly between Storage Nodes without considering the relative capacities of each node.
- Will not distribute erasure-coded data to Storage Nodes that are more than 70% full.

- Might decrease the performance of ILM operations and S3 and Swift client operations when it runs—additional resources are required to redistribute the erasure-coding fragments.
- Might need to be run more than once for systems with very large numbers of erasure-coded objects. To limit resource usage, the maximum number of moves is capped for each job.

When the EC rebalance procedure is complete:

- Erasure-coded data will have moved from Storage Nodes with less available space to Storage Nodes with more available space.
- The data protection of erasure-coded objects will be unchanged.
- Used (%) values might be different between Storage Nodes for two reasons:
  - Replicated object copies will continue to consume space on the existing nodes—the EC rebalance procedure does not move replicated data.
  - Larger-capacity nodes will be relatively less full than smaller-capacity nodes, even though all nodes will end up with approximately the same amount of erasure-coded data.

    For example, suppose three 200-TB nodes are each filled to 80% (200 × 0.8 = 160 TB on each node, or 480 TB for the site). If you add a 400-TB node and run the rebalance procedure, all nodes will now have approximately the same amount of erasure-code data (480/4 = 120 TB). However, the Used (%) for the larger node will be less than the Used (%) for the smaller nodes.
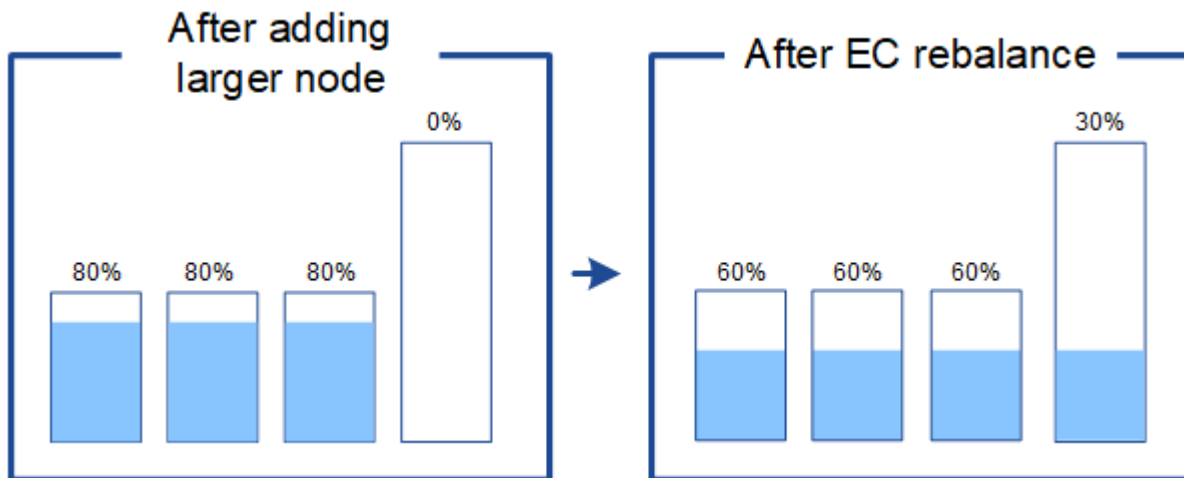
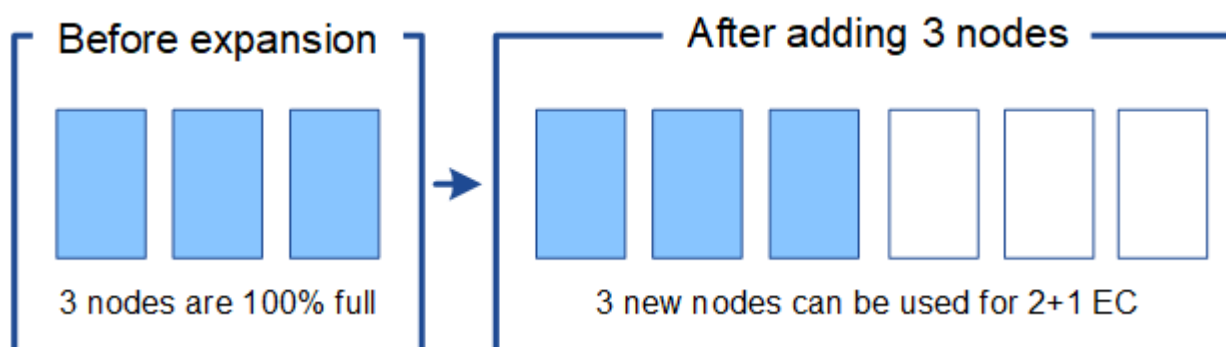**When to rebalance erasure-coded data**

Consider the following scenario:

- StorageGRID is running at a single site, which contains three Storage Nodes.
- The ILM policy uses a 2+1 erasure-coding rule for all objects larger than 1.0 MB and a 2-copy replication rule for smaller objects.
- All Storage Nodes have become completely full. The **Low Object Storage** alert has been triggered at the major severity level.

Before expansion

3 nodes are 100% full
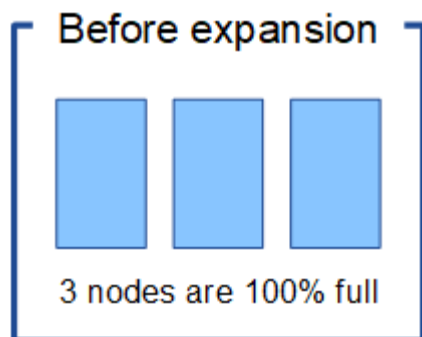
Rebalance is not required if you add enough nodes

To understand when EC rebalance is not required, suppose you added three (or more) new Storage Nodes. In this case, you don't need to perform EC rebalance. The original Storage Nodes will remain full, but new objects will now use the three new nodes for 2+1 erasure coding—the two data fragments and the one parity fragment can each be stored on a different node.

Before expansion

3 nodes are 100% full

After adding 3 nodes

3 new nodes can be used for 2+1 EC

> ⚠ While you can run the EC rebalance procedure in this case, moving the existing erasure-coded data will temporarily decrease the grid's performance, which might impact client operations.

Rebalance is required if you can't add enough nodes

To understand when EC rebalance is required, suppose you can only add two Storage Nodes, instead of three. Because the 2+1 scheme requires at least three Storage Nodes to have space available, the empty nodes can't be used for new erasure-coded data.

Before expansion

3 nodes are 100% full

After adding 2 nodes

2 new nodes cannot be used for EC
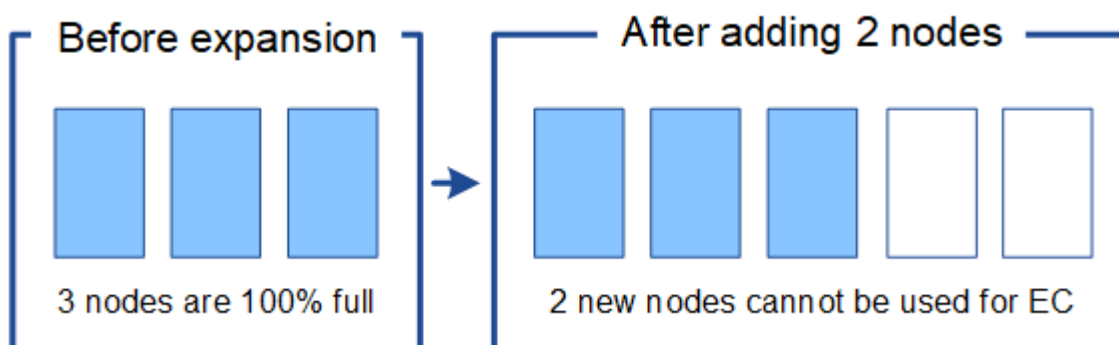
To make use of the new Storage Nodes, you should run the EC rebalance procedure. When this procedure runs, StorageGRID redistributes existing erasure-coded data and parity fragments among all Storage Nodes at the site. In this example, when the EC rebalance procedure is complete, all five nodes are now only 60% full,

and objects can continue to be ingested into the 2+1 erasure coding scheme on all Storage Nodes.

Before expansion	After adding 2 nodes	After EC rebalance
3 nodes are 100% full	2 new nodes cannot be used for EC	5 nodes are 60% full, EC can continue
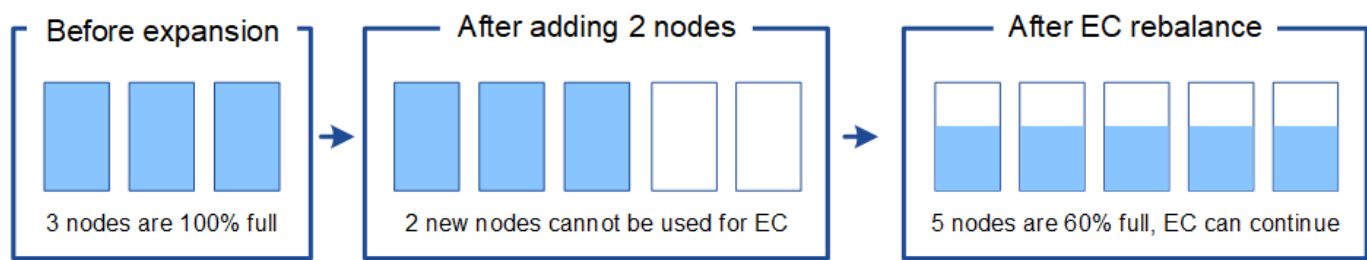
Recommendations for EC rebalancing

NetApp requires EC rebalancing if *all* of the following statements are true:

- You use erasure coding for your object data.
- The **Low Object Storage** alert has been triggered for one or more Storage Nodes at a site, indicating that the nodes are 80% or more full.
- You are unable to add enough new Storage Nodes for the erasure-coding scheme in use. See Add storage capacity for erasure-coded objects.
- Your S3 and Swift clients can tolerate lower performance for their write and read operations while the EC rebalance procedure is running.

You can optionally run the EC rebalance procedure if you prefer Storage Nodes to be filled to similar levels and your S3 and Swift clients can tolerate lower performance for their write and read operations while the EC rebalance procedure is running.

How EC rebalance procedure interacts with other maintenance tasks

You can't perform certain maintenance procedures at the same time you are running the EC rebalance procedure.

Procedure	Allowed during EC rebalance procedure?
Additional EC rebalance procedures	No. You can only run one EC rebalance procedure at a time.
Decommission procedure EC data repair job	No. - You are prevented from starting a decommission procedure or an EC data repair while the EC rebalance procedure is running. - You are prevented from starting the EC rebalance procedure while a Storage Node decommission procedure or an EC data repair is running.
Expansion procedure	No. If you need to add new Storage Nodes in an expansion, run the EC rebalance procedure after adding all new nodes.

Procedure	Allowed during EC rebalance procedure?
Upgrade procedure	No. If you need to upgrade StorageGRID software, perform the upgrade procedure before or after running the EC rebalance procedure. As required, you can terminate the EC rebalance procedure to perform a software upgrade.
Appliance node clone procedure	No. If you need to clone an appliance Storage Node, run the EC rebalance procedure after adding the new node.
Hotfix procedure	Yes. You can apply a StorageGRID hotfix while the EC rebalance procedure is running.
Other maintenance procedures	No. You must terminate the EC rebalance procedure before running other maintenance procedures.

How EC rebalance procedure interacts with ILM

While the EC rebalance procedure is running, avoid making ILM changes that might change the location of existing erasure-coded objects. For example, don't start using an ILM rule that has a different erasure coding profile. If you need to make such ILM changes, you should terminate the EC rebalance procedure.

Add metadata capacity

To ensure that adequate space is available for object metadata, you might need to perform an expansion procedure to add new Storage Nodes at each site.

StorageGRID reserves space for object metadata on volume 0 of each Storage Node. Three copies of all object metadata are maintained at each site, evenly distributed across all Storage Nodes.

You can use the Grid Manager to monitor the metadata capacity of Storage Nodes and to estimate how quickly metadata capacity is being consumed. In addition, the **Low metadata storage** alert is triggered for a Storage Node when the used metadata space reaches certain thresholds.

Note that a grid's object metadata capacity might be consumed faster than its object storage capacity, depending on how you use the grid. For example, if you typically ingest large numbers of small objects or add large quantities of user metadata or tags to objects, you might need to add Storage Nodes to increase metadata capacity even though sufficient object storage capacity remains.

For more information, see the following:

- Manage object metadata storage
- Monitor object metadata capacity for each Storage Node

Guidelines for increasing metadata capacity

Before adding Storage Nodes to increase metadata capacity, review the following guidelines and limitations:

- Assuming sufficient object storage capacity is available, having more space available for object metadata increases the number of objects you can store in your StorageGRID system.
- You can increase a grid's metadata capacity by adding one or more Storage Nodes to each site.
- The actual space reserved for object metadata on any given Storage Node depends on the Metadata Reserved Space storage option (system-wide setting), the amount of RAM allocated to the node, and the size of the node's volume 0.
- You can't increase metadata capacity by adding storage volumes to existing Storage Nodes, because metadata is stored only on volume 0.
- You can't increase metadata capacity by adding a new site.
- StorageGRID keeps three copies of all object metadata at every site. For this reason, the metadata capacity for your system is limited by the metadata capacity of your smallest site.
- When adding metadata capacity, you should add the same number of Storage Nodes to each site.

See the description of what Metadata Reserved Space is.

How metadata is redistributed when you add Storage Nodes

When you add Storage Nodes in an expansion, StorageGRID redistributes the existing object metadata to the new nodes at each site, which increases the overall metadata capacity of the grid. No user action is required.

The following figure shows how StorageGRID redistributes object metadata when you add Storage Nodes in an expansion. The left side of the figure represents volume 0 of three Storage Nodes before an expansion. Metadata is consuming a relatively large portion of each node's available metadata space, and the **Low metadata storage** alert has been triggered.

The right side of the figure shows how the existing metadata is redistributed after two Storage Nodes are added to the site. The amount of metadata on each node has decreased, the **Low metadata storage** alert is no longer triggered, and the space available for metadata has increased.
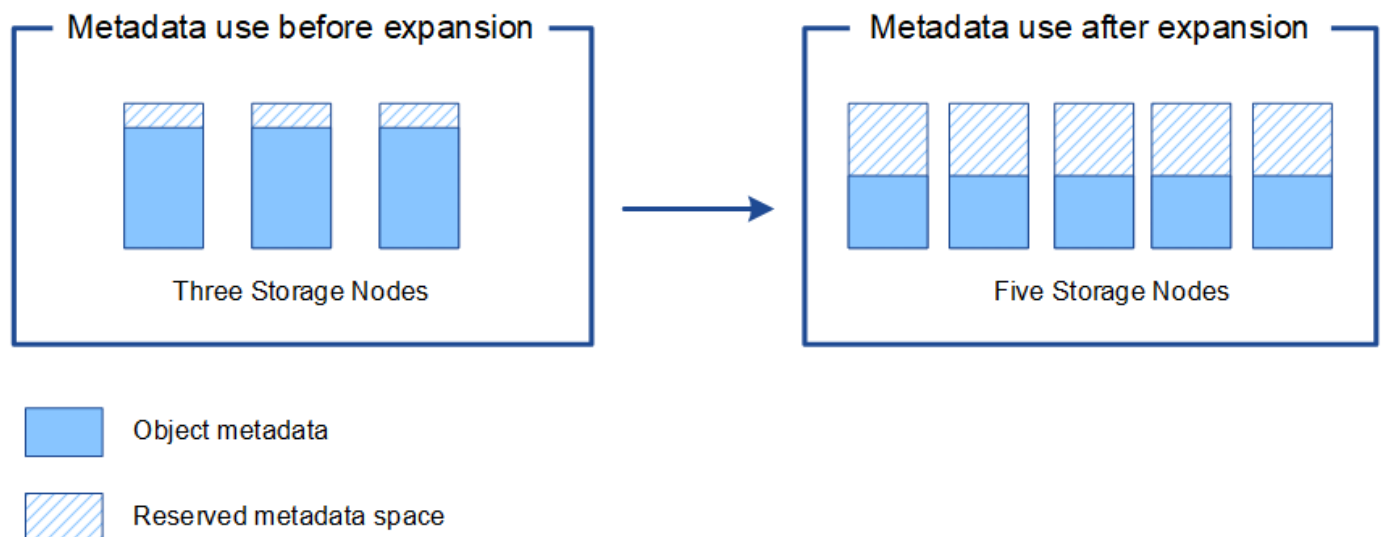
Add grid nodes to add capabilities to your system

You can add redundancy or additional capabilities to a StorageGRID system by adding new grid nodes to existing sites.

For example, you might choose to add Gateway Nodes to use in a high availability (HA) group, or you might add an Admin Node at a remote site to permit monitoring using a local node.

You can add one or more of the following types of nodes to one or more existing sites in a single expansion operation:

- Non-primary Admin Nodes
- Storage Nodes
- Gateway Nodes

When preparing to add grid nodes, be aware of the following limitations:

- The primary Admin Node is deployed during the initial installation. You can't add a primary Admin Node during an expansion.
- You can add Storage Nodes and other types of nodes in the same expansion.
- When adding Storage Nodes, you must carefully plan the number and location of the new nodes. See Guidelines for adding object capacity.
- If the **Set new node default** option is **Untrusted** on the Untrusted Client Networks tab on the Firewall control page, client applications that connect to expansion nodes using the Client Network must connect using a load balancer endpoint port (**CONFIGURATION** > **Security** > **Firewall control**). See the instructions to change the security setting for the new node and to configure load balancer endpoints.

Add a new site

You can expand your StorageGRID system by adding a new site.

Guidelines for adding a site

Before adding a site, review the following requirements and limitations:

- You can only add one site per expansion operation.
- You can't add grid nodes to an existing site as part of the same expansion.
- All sites must include at least three Storage Nodes.
- Adding a new site does not automatically increase the number of objects you can store. The total object capacity of a grid depends on the amount of available storage, the ILM policy, and the metadata capacity at each site.
- When sizing a new site, you must ensure that it includes enough metadata capacity.

  StorageGRID keeps a copy of all object metadata at every site. When you add a new site, you must ensure that it includes enough metadata capacity for the existing object metadata and enough metadata capacity for growth.

  For more information, see the following:

- Manage object metadata storage

- Monitor object metadata capacity for each Storage Node

- You must consider the available network bandwidth between sites, and the level of network latency. Metadata updates are continually replicated between sites even if all objects are stored only at the site where they are ingested.

- Because your StorageGRID system remains operational during the expansion, you must review ILM rules before starting the expansion procedure. You must ensure that object copies aren't stored to the new site until the expansion procedure is complete.

  For example, before you begin the expansion, determine if any rules use the default storage pool (All Storage Nodes). If they do, you must create a new storage pool that contains the existing Storage Nodes and update your ILM rules to use the new storage pool. Otherwise, objects will be copied to the new site as soon as the first node at that site becomes active.

  For more information about changing ILM when adding a new site, see the example for changing an ILM policy.