



# 对象的存储方式（复制或纠删编码） StorageGRID 11.7

NetApp  
April 12, 2024

# 目录

- 对象的存储方式（复制或纠删编码） ..... 1
  - 什么是复制？ ..... 1
  - 为什么不应使用单副本复制 ..... 2
  - 什么是纠删编码？ ..... 4
  - 什么是纠删编码方案？ ..... 6
  - 纠删编码的优势，劣势和要求 ..... 8

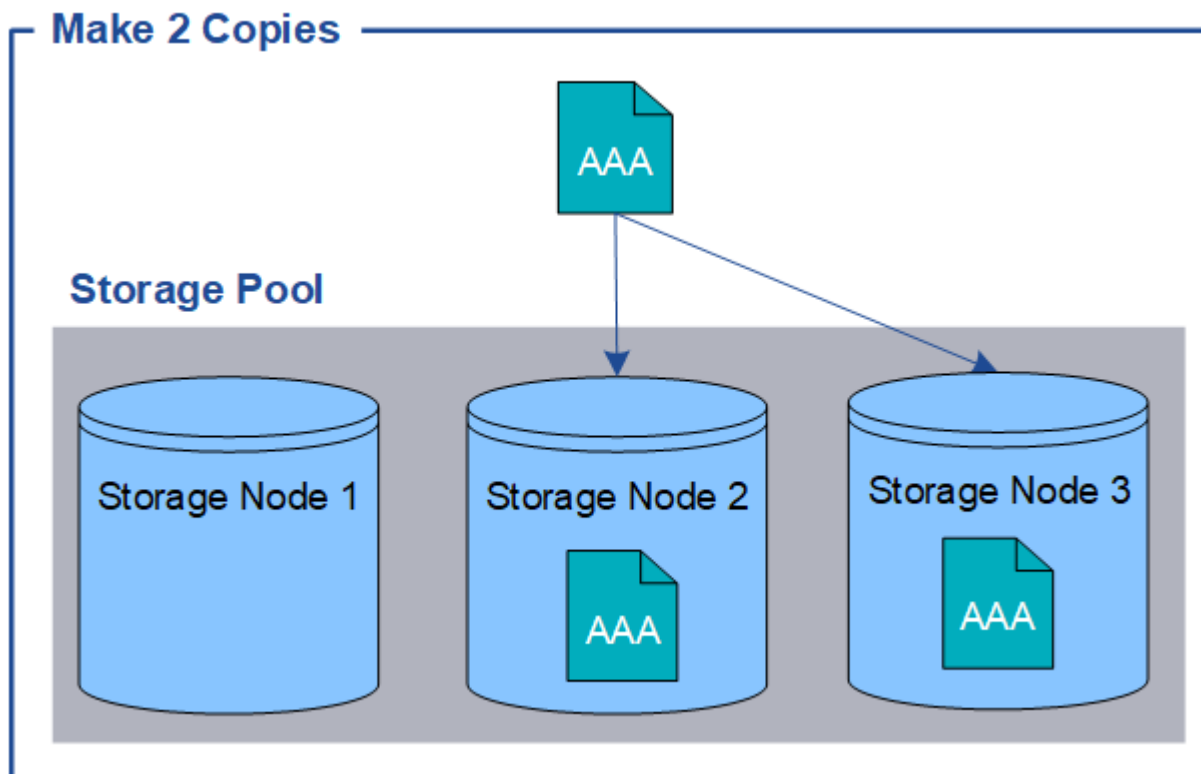
# 对象的存储方式（复制或纠删编码）

## 什么是复制？

复制是 StorageGRID 用于存储对象数据的两种方法之一。当对象与使用复制的 ILM 规则匹配时，系统会为对象数据创建精确副本，并将这些副本存储在存储节点或归档节点上。

在配置创建复制副本的 ILM 规则时，您可以指定应创建多少个副本，这些副本应放置在何处以及应将这些副本存储在每个位置的时间长度。

在以下示例中，ILM 规则指定将每个对象的两个复制副本放置在包含三个存储节点的存储池中。



当 StorageGRID 将对象与此规则匹配时，它会为该对象创建两个副本，并将每个副本放置在存储池中的不同存储节点上。这两个副本可以放置在三个可用存储节点中的任意两个上。在这种情况下，规则会将对象副本放置在存储节点 2 和 3 上。由于有两个副本，因此，如果存储池中的任何节点出现故障，可以检索此对象。



StorageGRID 只能在任何给定存储节点上存储一个对象的一个复制副本。如果您的网格包含三个存储节点，并且您创建了一个 4 副本 ILM 规则，则只会创建三个副本—每个存储节点一个副本。系统将触发 \* 无法实现 ILM 放置 \* 警报，以指示无法完全应用 ILM 规则。

### 相关信息

- ["什么是纠删编码？"](#)
- ["什么是存储池？"](#)
- ["通过复制和纠删编码实现站点丢失保护"](#)

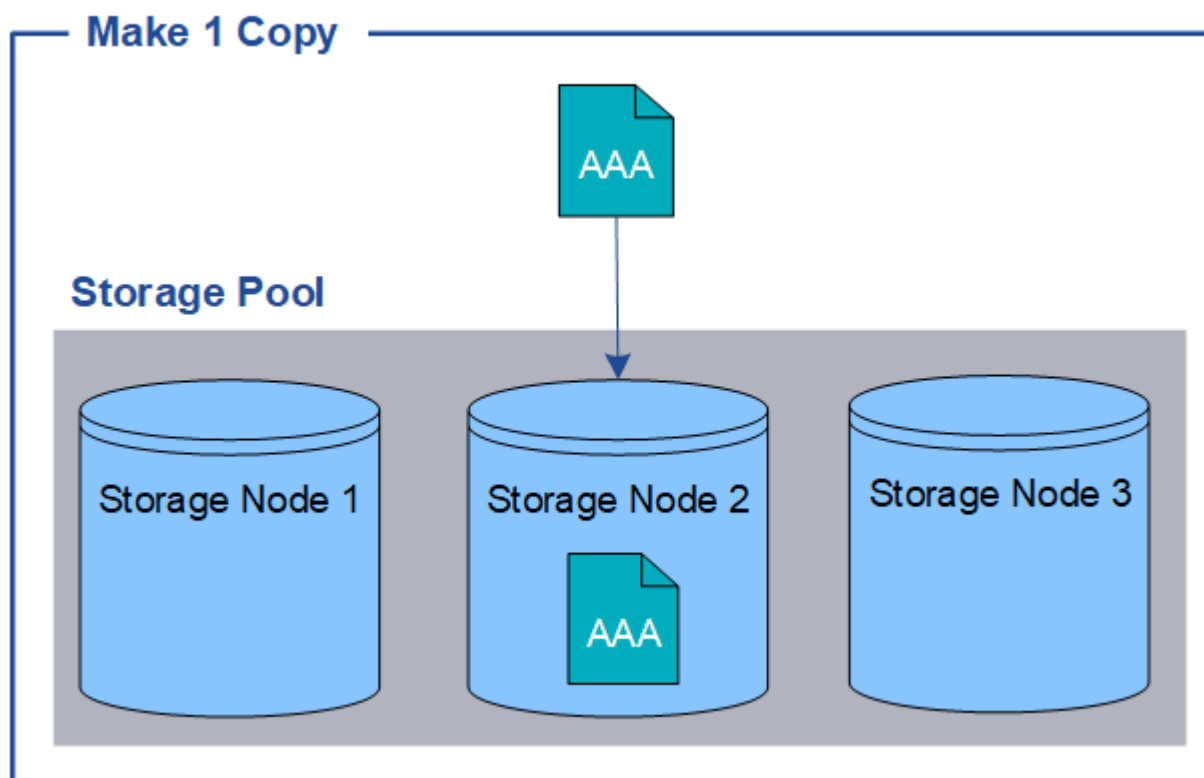
## 为什么不应使用单副本复制

在创建 ILM 规则以创建复制副本时，您应始终在放置说明中指定至少两个副本，以便在任意时间段内使用。



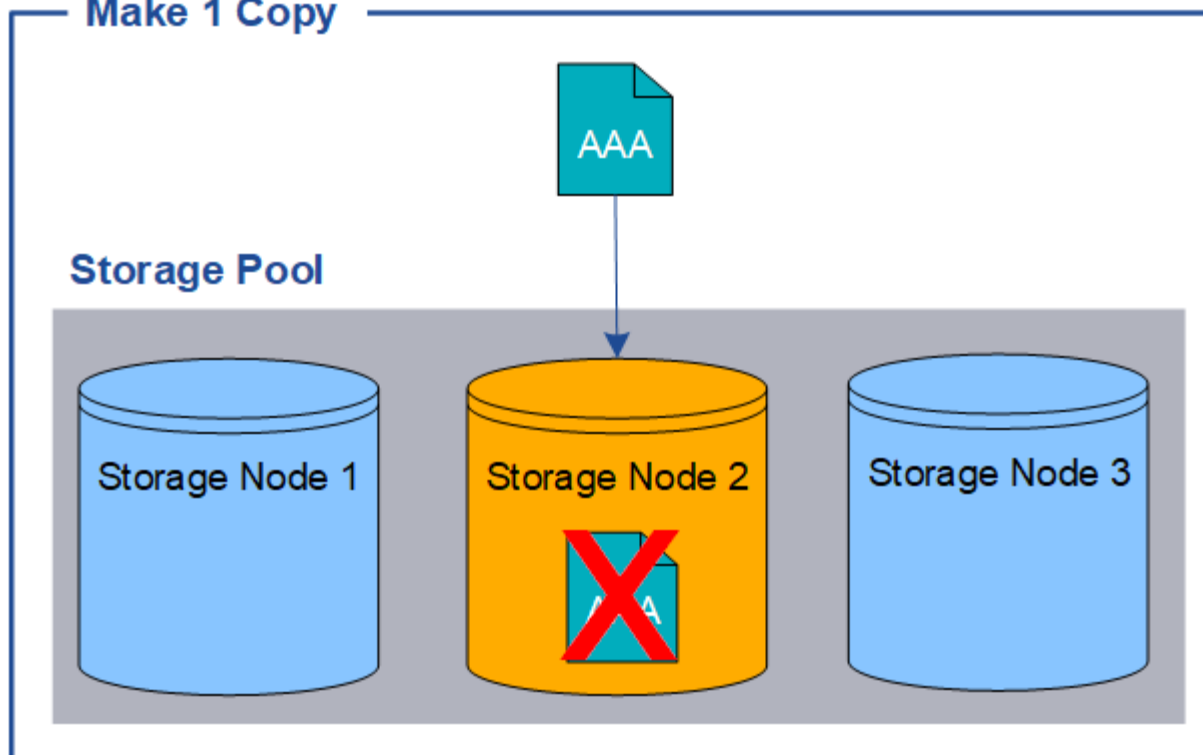
请勿使用在任何时间段内仅创建一个复制副本的 ILM 规则。如果某个对象只存在一个复制副本，则在存储节点出现故障或出现严重错误时，该对象将丢失。在升级等维护过程中，您还会暂时失去对对象的访问权限。

在以下示例中，Make 1 Copy ILM 规则指定将对象的一个复制副本放置在包含三个存储节点的存储池中。如果载入的对象与此规则匹配，则 StorageGRID 仅会在一个存储节点上放置一个副本。



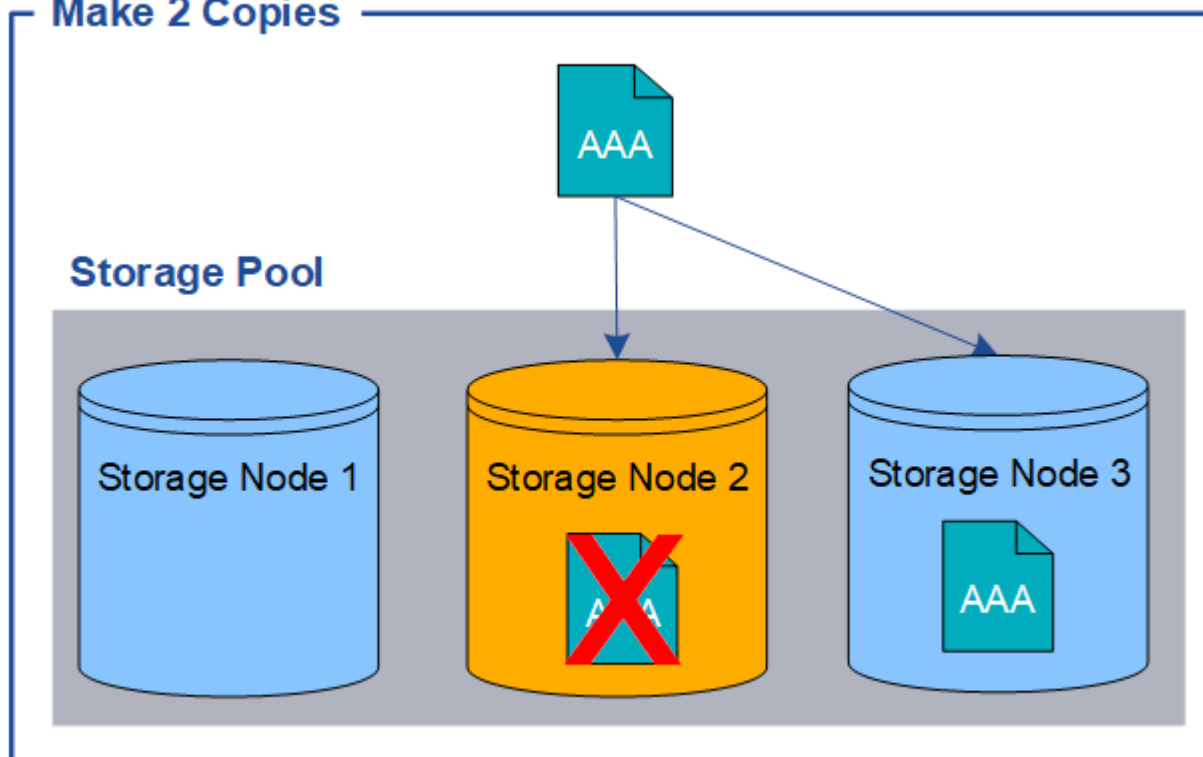
如果 ILM 规则仅创建一个对象的一个复制副本，则在存储节点不可用时，此对象将无法访问。在此示例中，只要存储节点 2 脱机，例如在升级或其他维护操作步骤 期间，您将暂时无法访问对象 AAA。如果存储节点 2 发生故障，您将完全丢失对象 AAA。

## Make 1 Copy



为了避免丢失对象数据，您应始终为要通过复制保护的所有对象创建至少两个副本。如果存在两个或更多副本，则在一个存储节点出现故障或脱机时，您仍可以访问此对象。

## Make 2 Copies



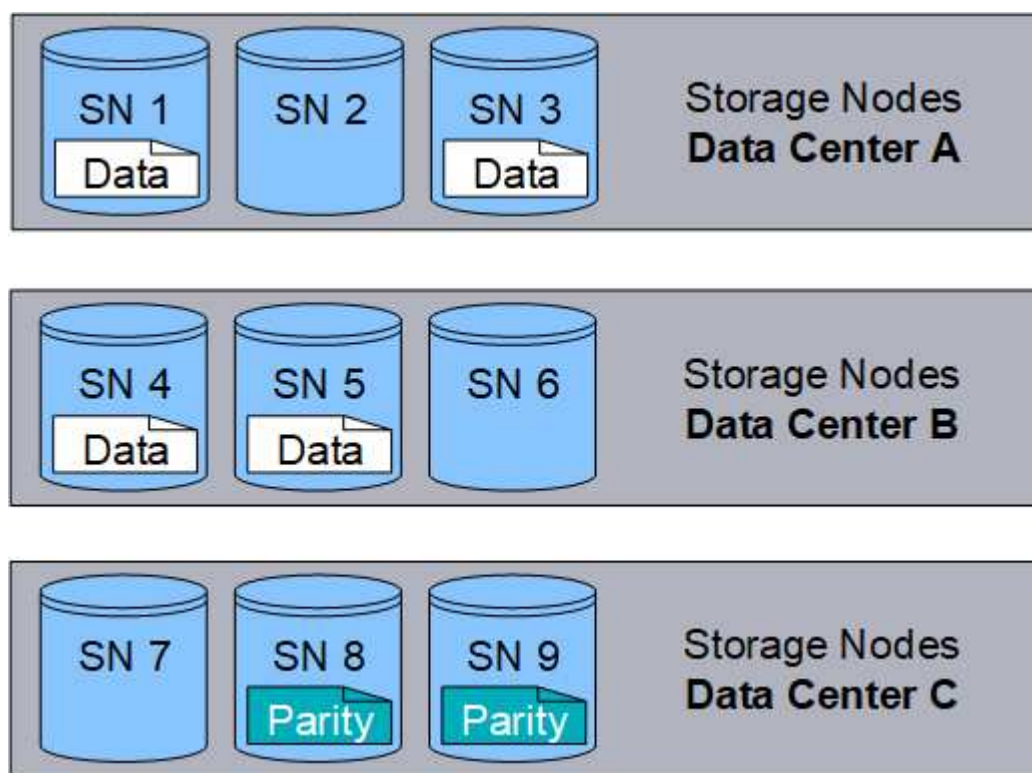
## 什么是纠删编码？

纠删编码是StorageGRID 用于存储对象数据的两种方法之一。如果对象与使用纠删编码的ILM规则匹配、则这些对象会划分为数据片段、并计算额外的奇偶校验片段、每个片段会存储在不同的存储节点上。

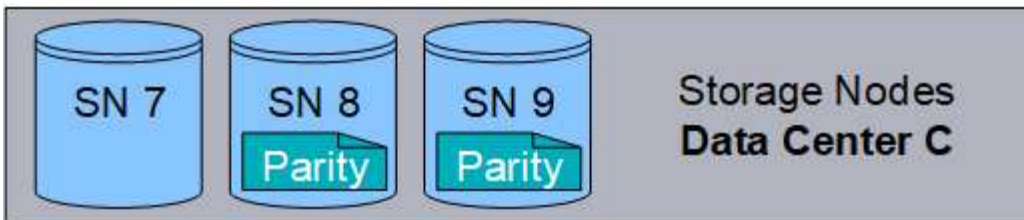
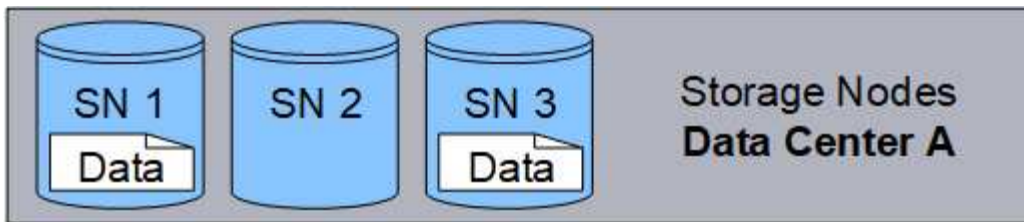
访问某个对象时，系统会使用存储的片段重新组合该对象。如果数据或奇偶校验片段损坏或丢失，则纠删编码算法可以使用剩余数据和奇偶校验片段的子集重新创建该片段。

创建ILM规则时、StorageGRID 会创建支持这些规则的纠删编码配置文件。您可以查看纠删编码配置文件列表、["重命名纠删编码配置文件"](#)或 ["如果纠删编码配置文件当前未在任何ILM规则中使用、则停用该配置文件"](#)。

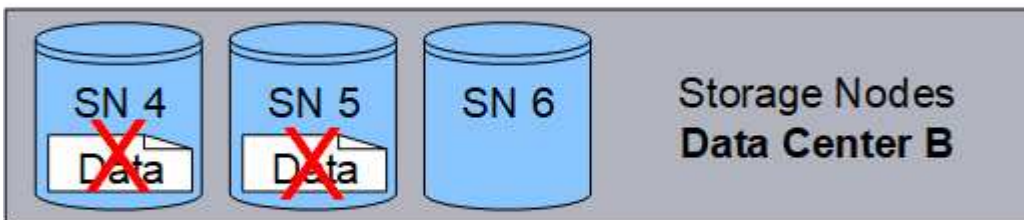
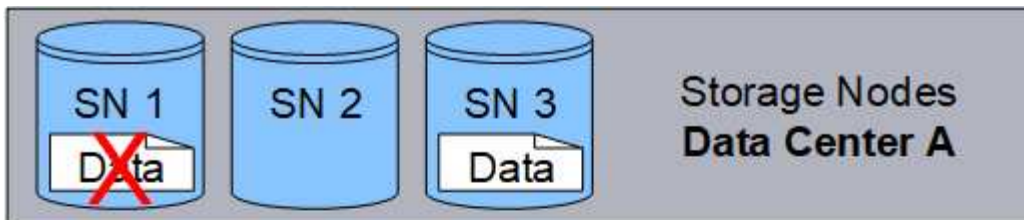
以下示例说明了如何对对象数据使用纠删编码算法。在此示例中， ILM 规则使用 4+2 纠删编码方案。每个对象都会被划分为四个相等的数据片段，并根据对象数据计算两个奇偶校验片段。六个片段中的每个片段都存储在三个数据中心站点的不同节点上，以便为节点故障或站点丢失提供数据保护。



4+2纠删编码方案可以通过多种方式进行配置。例如、您可以配置一个包含六个存储节点的单站点存储池。适用于 ["站点丢失保护"](#)，则可以使用包含三个站点的存储池，每个站点上有三个存储节点。只要六个片段中的任意四个（数据或奇偶校验）仍然可用，就可以检索对象。最多可以丢失两个片段，而不会丢失对象数据。如果整个站点丢失、只要所有其他碎片仍可访问、就可以检索或修复对象。



如果丢失两个以上的存储节点，则无法检索此对象。



相关信息

- "什么是复制？"
- "什么是存储池？"
- "什么是纠删编码方案？"

- ["重命名纠删编码配置文件"](#)
- ["停用纠删编码配置文件"](#)

## 什么是纠删编码方案？

纠删编码方案可控制为每个对象创建的数据片段数量和奇偶校验片段数量。

在为 ILM 规则配置纠删编码配置文件时、您可以根据计划使用的存储池中的存储节点和站点数量选择可用的纠删编码方案。

StorageGRID 系统使用 Reed-Solomon 纠删编码算法。算法会将对象切片到  $k$  数据片段和数据的统计数据  $m$  奇偶校验片段。  $k + m = n$  片段分布在各处  $n$  用于提供数据保护的存储节点。一个对象最多可支持  $m$  碎片丢失或损坏。检索或修复对象，  $k$  需要片段。

如果要为创建经过删除编码的副本的规则选择要使用的存储池、请对存储池使用以下准则：

- 存储池必须包含三个或更多站点，或者只包含一个站点。



如果存储池包含两个站点、则不能使用纠删编码。

- [包含三个或更多站点的存储池的纠删编码方案](#)
- [单站点存储池的纠删编码方案](#)
- 请勿使用包含默认站点"所有站点"的存储池。
- 存储池应至少包含  $k+m + 1$  存储节点。

所需的最小存储节点数为  $k+m$ 。但是，如果所需的存储节点暂时不可用，则至少添加一个存储节点有助于防止载入失败或 ILM 回退。

纠删编码方案的存储开销是通过除奇偶校验片段数计算得出的  $(m) (k)$ 。您可以使用存储开销计算每个擦除编码对象所需的磁盘空间量：

$$\text{disk space} = \text{object size} + (\text{object size} * \text{storage overhead})$$

例如，如果使用 4+2 方案存储一个 10 MB 的对象（存储开销为 50%），则该对象将占用 15 MB 的网格存储。如果使用 6+2 方案存储同一个 10 MB 对象（存储开销为 33%），则该对象将占用大约 13.3 MB 的空间。

选择总值最低的纠删编码方案  $k+m$  满足您的需求。片段数量较少的纠删编码方案在计算上总体上更高效、因为每个对象创建和分布(或检索)的片段较少、片段大小较大、性能较高、并且在需要更多存储时、扩展中添加的节点较少。(有关规划存储扩展的信息、请参见 ["扩展 StorageGRID 的说明"](#))

### 包含三个或更多站点的存储池的纠删编码方案

下表介绍了 StorageGRID 当前支持的纠删编码方案，该方案适用于包含三个或更多站点的存储池。所有这些方案都提供站点丢失保护。一个站点可能会丢失，但对象仍可访问。

对于可提供站点丢失保护的纠删编码方案、存储池中建议的存储节点数超过  $k+m + 1$  因为每个站点至少需要三个存储节点。



纠删编码方案 ( $k+m$ )	已部署站点的最小数量	每个站点的建议存储节点数	建议的存储节点总数	站点丢失保护?	存储开销
4+2	3.	3.	9	是的。	50%
6+2	4.	3.	12	是的。	33%
8+2	5.	3.	15	是的。	25%
6+3.	3.	4.	12	是的。	50%
9+3.	4.	4.	16.	是的。	33%
2+1	3.	3.	9	是的。	50%
4+1	5.	3.	15	是的。	25%
6+1	7.	3.	21	是的。	17%
7+5.	3.	5.	15	是的。	71%



StorageGRID 要求每个站点至少有三个存储节点。要使用 7+5 方案，每个站点至少需要四个存储节点。建议每个站点使用五个存储节点。

在选择提供站点保护的纠删编码方案时，请平衡以下因素的相对重要性：

- \* 碎片数量 \*：当碎片总数减少时，性能和扩展灵活性通常会提高。
- 容错：容错功能通过增加奇偶校验段(即、何时  $m$  具有更高的值。)
- 网络流量：从故障中恢复时、使用包含更多片段的方案(即、的总流量较高)  $k+m$ )创建更多网络流量。
- \* 存储开销 \*：开销较高的方案需要每个对象更多的存储空间。

例如，在选择 4+2 方案和 6+3 方案（两者都有 50% 的存储开销）时，如果需要额外的容错功能，请选择 6+3 方案。如果网络资源受限，请选择 4+2 方案。如果所有其他因素相等，请选择 4+2，因为其碎片总数较低。



如果您不确定要使用的方案，请选择 4+2 或 6+3，或者联系技术支持。

## 单站点存储池的纠删编码方案

单站点存储池支持为三个或更多站点定义的所有纠删编码方案，但前提是该站点具有足够的存储节点。

所需的最小存储节点数为  $k+m$ 、但具有的存储池  $k+m+1$  建议使用存储节点。例如，2+1 纠删编码方案要求一个存储池至少包含三个存储节点，但建议使用四个存储节点。

纠删编码方案（ $k+m$ ）	存储节点的最小数量	建议的存储节点数	存储开销
4+2	6.	7.	50%
6+2	8.	9	33%
8+2	10	11.	25%
6+3.	9	10	50%
9+3.	12	13	33%
2+1	3.	4.	50%
4+1	5.	6.	25%
6+1	7.	8.	17%
7+5.	12	13	71%

## 纠删编码的优势，劣势和要求

在决定是使用复制还是纠删编码来保护对象数据不会丢失之前，您应了解纠删编码的优点，缺点和要求。

### 纠删编码的优势

与复制相比，纠删编码可提高可靠性，可用性和存储效率。

- **\* 可靠性 \***：可靠性通过容错来衡量—即，在不丢失数据的情况下可以同时发生的故障数量。通过复制，多个相同的副本会存储在不同的节点上以及不同的站点上。通过纠删编码，对象会编码为数据和奇偶校验片段，并分布在多个节点和站点上。这种分散方式可同时提供站点和节点故障保护。与复制相比，纠删编码可提高可靠性，而存储成本相当。
- **\* 可用性 \***：可用性可定义为在存储节点出现故障或无法访问时检索对象的功能。与复制相比，纠删编码可以以相当的存储成本提高可用性。
- **\* 存储效率 \***：对于相似级别的可用性和可靠性，通过纠删编码保护的對象比通过复制保护的相同对象占用的磁盘空间更少。例如，复制到两个站点的 10 MB 对象会占用 20 MB 的磁盘空间（两个副本），而采用 6+3 纠删编码方案在三个站点之间进行纠删编码的对象只会占用 15 MB 的磁盘空间。



擦除编码对象的磁盘空间计算为对象大小加上存储开销。存储开销百分比是奇偶校验片段数除以数据片段数。

### 纠删编码的缺点

与复制相比，纠删编码具有以下缺点：

- 建议增加存储节点和站点的数量、具体取决于纠删编码方案。相比之下、如果复制对象数据、则每个副本只需要一个存储节点。请参见 ["包含三个或更多站点的存储池的纠删编码方案"](#) 和 ["用于单站点存储池的纠删编码方案"](#)。
- 存储扩展的成本和复杂性增加。要扩展使用复制的部署、您需要在创建对象副本的每个位置添加存储容量。要扩展使用纠删编码的部署，您必须同时考虑使用的纠删编码方案以及现有存储节点的容量。例如、如果等待现有节点达到100%全满、则必须至少添加  $k+m$  存储节点、但是、如果在现有节点已用容量达到70%时进行扩展、则可以为每个站点添加两个节点、同时仍然可以最大程度地提高可用存储容量。有关详细信息，请参见 ["为经过纠删编码的对象添加存储容量"](#)。
- 在分布在不同地理位置的站点之间使用纠删编码时，检索延迟会增加。与在本地复制并提供的对象（客户端连接的同一站点）相比，通过 WAN 连接检索经过纠删编码并分布在远程站点上的对象的对象片段所需时间更长。
- 在地理位置分散的站点之间使用纠删编码时，检索和修复的 WAN 网络流量使用率较高，尤其是频繁检索的对象或通过 WAN 网络连接进行对象修复。
- 当您在站点间使用纠删编码时，最大对象吞吐量会随着站点间网络延迟的增加而急剧下降。这一减少是由于 TCP 网络吞吐量相应减少，从而影响 StorageGRID 系统存储和检索对象片段的速度。
- 提高计算资源的利用率。

## 何时使用纠删编码

纠删编码最适合以下要求：

- 大于 1 MB 的对象。



纠删编码最适合大于 1 MB 的对象。不要对小于200 KB的对象使用纠删编码、以避免管理非常小的经过纠删编码的片段所产生的开销。

- 长期或冷存储，用于存储不经常检索的内容。
- 高数据可用性和可靠性。
- 防止发生完整的站点和节点故障。
- 存储效率。
- 需要高效数据保护的单站点部署，只需一个纠删编码副本，而不是多个复制副本。
- 站点间延迟小于 100 毫秒的多站点部署。

## 版权信息

版权所有 © 2024 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

## 商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。