

重复数据删除白皮书

基于 Active Backup for Business 2.3.0



目录

简介	3
数据缩减技术	4
重复数据删除过程	
更改块跟踪 (CBT)	
重复数据删除率	
结论	8

查找您所需的信息

Synology 发布了多种支持文档。

在[知识中心](#)内，您可以找到有用的帮助和常见问题文章，以及将过程划分成便捷步骤的视频教程。您还可以找到用户指南、解决方案指南、手册和白皮书。经验丰富的用户和管理员可在技术性管理员指南和开发人员指南中找到答案和指导。

您是否遇到了问题，并且无法在我们的官方文档中找到解决方案？可在[Synology Community](#)中搜索用户和支持人员提供的数百个答案，或者通过 Web 表单、电子邮件或电话联系 [Synology 技术支持](#)。



简介

无论您使用何种存储设备或系统，数据丢失或系统故障都可能造成灾难性后果。对您最重要的设备（例如工作计算机或服务器）应该采取一种具体而有效的备份策略，以帮助您在发生自然灾害或其他技术问题避免意外的系统故障。

Active Backup for Business (ABB) 使您可从 Synology NAS 的单一中央控制台管理所有 VMware、Hyper-V、Windows、Linux 和文件服务器备份。借助用于 Windows 设备、虚拟机和单个文件的即时还原选项，您可以使关键服务保持正常运行，即使在发生突发灾难时也是如此。除此之外，Active Backup for Business 还利用了一些最新技术（如**重复数据删除**和**增量备份**）来增强备份性能，并尽可能减少将备份数据所需的物理空间量。由于这些方法在不同的备份服务中十分常用，因此我们的用户经常会询问 ABB 重复数据删除率所采用的技术和测量方法。

本文档旨在介绍一些使用的重复数据删除技术，并定义在 Synology Active Backup for Business 中实施的重复数据删除率。

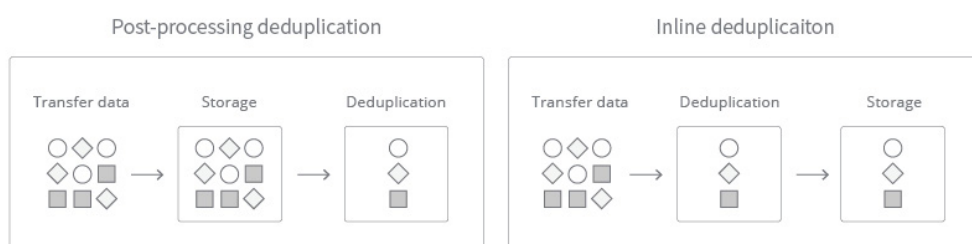
数据缩减技术

数据缩减技术旨在减少备份数据占用的存储空间，从而优化存储使用并让用户可以备份尽可能多的设备。Active Backup for Business 实施了先进的技术来进行数据缩减并提高总体备份性能。

重复数据删除过程

重复数据删除是通过移除任何冗余数据来优化存储空间的有效方法。重复数据删除过程可以分为两种类型：“内联”和“后期处理”。**内联重复数据删除**会在数据写入存储之前移除所有冗余数据，而**后期处理重复数据删除**仅在备份过程完成之后移除数据。

根据备份解决方案，可同时在备份代理程序和备份目的地上实施内联重复数据删除。但是，当通过代理程序实施重复数据删除时，数据扫描、重复数据删除和传输都会同时进行，从而导致备份过程更加耗时且影响性能（备份目的地）。



Synology Active Backup for Business 在其设计中融入了**内联重复数据删除**功能，但重复数据删除过程仅在写入存储之前在目的地进行。这样做是为了避免在备份来源上过度消耗时间和 CPU 资源，这通常是由来源端重复数据删除造成的。此外，Active Backup for Business 允许用户将各种系统和设备备份到可能包含冗余数据的单个目的地。

为了跨多个目标消除数据的所有冗余副本，Active Backup for Business 还实施了**全局重复数据删除**。全局重复数据删除使 Active Backup for Business 可以在单个共享文件夹中跨多个备份任务进行重复数据删除，从而优化实际存储消耗，而不会影响总体备份性能。

关于重复数据删除机制，Active Backup for Business 使用唯一标识符对备份的每个文件和文件块进行分类。Active Backup for Business 重复数据删除引擎通过将备份数据划分为各个 4KB 文件块并为每个文件块分配唯一的指纹来实现此目的。此指纹由重复数据删除引擎进行计算，该引擎使用 **SHA-256 算法** 标识每个数据块。

为文件块创建指纹后，系统会为每个存储的文件块创建指纹索引。此索引用于将传入文件块与其存储的对应文件块进行匹配。一旦确定，重复数据删除引擎便会立即移除冗余数据，从而优化设备的存储使用。

更改块跟踪 (CBT)

更改块跟踪 (CBT) 是一种增量备份技术，常用于多种不同的备份解决方案。增量备份是一种备份类型，仅备份自最新备份版本以来修改或新建的数据。CBT 可以通过仅备份更改的数据块（而不是每次完成完整备份）来节省时间和资源，从而帮助进行增量备份。因此，通过备份应用程序传输的数据大小（即备份数据传输大小）最终会小于备份来源上的原始数据大小。

Active Backup for Business 在其架构中包含了此功能，用于备份到计算机、物理服务器和虚拟机。ABB 会为备份任务执行永久增量备份以尽可能增加可用备份版本的数量，并通过在执行完整备份后仅备份修改的数据来尽可能减少备份保留所用的存储容量。



重复数据删除率

定义

重复数据删除率是用于评估备份解决方案重复数据删除性能有效性的指标。对于选择备份解决方案的用户而言，此比率通常是一个决定性因素，因为它使用户可以分析不同提供商间的不同功能和性能。但是，重复数据删除率可能会因使用的重复数据删除技术和数据的一般特征而异。

重复数据删除率或压缩比定义为要存储的数据量与在移除任何冗余数据后用于在备份目的地存储备份数据的实际存储空间量的度量。

例如：您从设备备份了 100 GB 数据，备份存储服务器上使用的实际空间量为 20 GB。将备份数据量 (100 GB) 除以备份存储服务器上使用的空间 (20 GB) 会得到重复数据删除率为 5:1，或重复数据删除率为 80%。

重复数据删除比率也在市场营销广泛用于宣传备份产品的优势。例如，促销活动可能会说，备份产品的重复数据删除率比市场上其他产品高 20 倍。但是，重复数据删除率会因使用的重复数据删除技术、数据的一般特征、使用的备份方法以及重复数据删除率的计算方式而异。

因而重复数据删除率可能并不总是比较不同产品的最准确方法。因此，请务必记住每个产品的重复数据删除率计算方式，以及得到的备份目的地实际占用的存储量。

计算

重复数据删除率可以使用以下值进行计算：

- 移除重复数据之前的备份数据总容量
- 备份数据传输大小
- 备份目的地上使用的实际容量

通常，计算该比率的一个好方法是将**备份数据传输大小**除以**备份目的地上使用的实际容量**。这是 Synology 使用的方法。

另一方面，一些备份提供商会将**移除重复数据之前的备份数据总容量**除以**备份目的地上使用的实际容量**，这可能并不一定会生成准确的重复数据删除率。

例如，请考虑以下情形：

- 备份解决方案：A
- 移除重复数据之前的备份数据总容量：6 TB
- 备份数据传输大小：1.2 TB
- 备份目的地上使用的实际容量：0.83 TB
- 宣传的重复数据删除率：7.2:1

在此情形中，宣传的备份解决方案 A 重复数据删除率为 **7.2:1** (6 TB/0.83 TB)。但是，此计算并不完全准确。这是因为只传输了 **1.2 TB 的备份数据**，冗余数据量随后通过 CBT 进行缩减，数据进一步压缩到 **0.83 TB** 以便存储在备份目的地上。

备份解决方案 A 将**备份数据的总容量**除以**备份目的地上使用的实际容量**以获取更高的重复数据删除率。但实际上，只有**备份数据传输大小**显示将进行重复数据删除的实际数据量。

将**备份数据传输大小**除以**备份目的地上使用的实际容量**会生成更准确的结果。因此，重复数据删除计算可以按以下所示进行调整：

重复数据删除率

备份数据传输大小	÷	备份目的地上使用的实际容量	=	重复数据删除率
1.2 TB		0.83 TB		1.44:1

重复数据删除率

(备份数据传输大小	-	备份目的地上使用的实际容量)	÷	备份数据传输大小	x 100%	=	重复数据删除率
	1.2 TB		0.83 TB			1.2 TB			30.8%

备份方法

可能影响重复数据删除率的另一个因素是**备份方法**，特别是备份是否全部为**完整备份**或是否使用**增量备份**。仅执行完整备份的备份系统似乎比使用增量备份的备份具有更好的重复数据删除率。这是因为增量备份仅传输更改数据块，并在传输数据之前移除任何冗余数据，从而减少备份目的地上使用的实际容量。另一方面，与原始或以前的备份版本相比，完整备份包含大量冗余数据。

总之，尽管增量备份解决方案表现出的重复数据删除率可能不如完整备份解决方案，但它们在备份目的地上仍会使用较少的实际存储空间。

Active Backup for Business 重复数据删除率

如前所述，使用**移除重复数据之前的备份数据总容量**计算重复数据删除率可能会生成更好的重复数据删除率，但它们可能并未反映进行重复数据删除的实际数据量。Active Backup for Business 改为利用**备份数据传输大小**，因为这代表了将进行重复数据删除的实际数据量。如下表所示，Active Backup for Business 的重复数据删除技术使得**备份目的地上使用的实际容量**更少，这证明它总体上为用户节省了更多的存储空间。

总之，用户可以更加关注与备份传输大小相关的实际占用存储量，以便更好地了解该产品如何帮助您节省存储空间，而不是依赖于产品的重复数据删除率。

下表提供 **Active Backup for Business** 的实际重复数据删除率，并将其与其他备份提供商的重复数据删除率进行比较。

提供者	备份数据传输大小	备份目的地上使用的实际容量	重复数据删除率	重复数据删除率
Synology Active Backup for Business	1.2 TB	637.2 GB	1.2 TB/0.62 TB = 1.93:1	48.18%
提供商 A	1.2 TB	819.8 GB	1.2 TB/0.8 TB = 1.5:1	33.33%
提供商 B	1.2 TB	847.7 GB	1.2 TB/0.82 TB = 1.45:1	31.08%

备份来源内容：Windows 计算机 x 9，虚拟机上的 Windows Server x 1



结论

重复数据删除是一个复杂过程，它允许用户减少冗余数据量，以便更高效地管理备份和优化存储空间。由于重复数据删除是在一些不同备份解决方案中使用的常见技术，因此许多用户希望了解重复数据删除率，以比较不同的产品。但是，由于有多种测量方法可用，并且有不同的因素可能会影响重复数据删除测量技术，因此每个产品提供的比率可能并不总是可靠。因而在考虑不同产品的重复数据删除率时，更重要的是关注进行重复数据删除之后用于存储备份数据的实际物理存储空间量。

Synology Active Backup for Business 不仅提供了可在各种平台上使用的全面备份解决方案，还提供了一个可实现快速、可靠和有效重复数据删除的架构，让您可以节省时间并充分利用存储空间。



**SYNOLOGY
INC.**

新北市板桥区
远东路1号9楼·邮编 220545
台湾
电话：+886 2 2955 1814

**SYNOLOGY
AMERICA CORP.**

3535 Factoria Blvd SE, Suite #200,
Bellevue, WA 98006
USA
电话：+1 425 818 1587

**SYNOLOGY
UK LTD.**

Unit 5 Danbury Court, Linford
Wood, Milton Keynes, MK14 6PL,
英国
电话：+44 (0)1908048029

**SYNOLOGY
FRANCE SARL**

102 Terrasse Boieldieu (TOUR W)
92800 Puteaux
France
电话：+33 147 176288

**SYNOLOGY
GMBH**

Grafenberger Allee 295
40237 Düsseldorf
Deutschland
电话：+49 211 9666 9666

**群晖网络科技
(上海)有限公司**

中国上海
静安区天目西路511号201
室
中国

**SYNOLOGY
JAPAN CO., LTD.**

4F, No. 3-1-2, Higashikanda
Chiyoda-ku, Tokyo, 101-0031
Japan

Synology®



synology.com

Synology 可随时修改产品规格与说明·恕不另行通知·版权所有 © 2022 Synology Inc. 保留所有权利·®
Synology 及其他 Synology 产品名称均为 Synology Inc. 所有的专有标志或注册商标·本文档提及的其他产品名称
及公司名称为其各自所有人的商标·