

重複資料刪除和壓縮

技術白皮書

2025 年四月

公告

版權

© 版權所有 2025 QSAN 廣盛科技保留所有權利。未經 QSAN 廣盛科技書面許可，不得複製或傳播本文件的任何部分。

QSAN 認為本出版物在發布之日內容準確無誤。資訊如有更改，恕不另行通知。

商標

- QSAN、QSAN 標誌、QSAN.com 是 QSAN 廣盛科技的商標或註冊商標。
- Microsoft、Windows、Windows Server 和 Hyper-V 是 Microsoft Corporation 在美國和 / 或其他國家 / 地區的商標或註冊商標。
- Linux 是 Linus Torvalds 在美國和 / 或其他國家 / 地區的商標。
- UNIX 是 The Open Group 在美國和其他國家 / 地區的註冊商標。
- Mac 和 OS X 是 Apple Inc. 在美國和其他國家 / 地區的註冊商標。
- Java 和所有基於 Java 的商標和標誌是 Oracle 和 / 或其附屬公司的商標或註冊商標。
- VMware、ESXi 和 vSphere 是 VMware, Inc. 在美國和 / 或其他國家 / 地區的註冊商標或商標。
- Citrix 和 Xen 是 Citrix Systems, Inc. 在美國和 / 或其他國家 / 地區的註冊商標或商標。
- 本文件中用於宣稱擁有商標和名稱的實體或其產品的其他商標和商品名稱均為其各自所有者的財產。

目錄

公告	i
注意	vi
前言	vii
技術支援	vii
資訊、提示和注意事項	vii
1. 概述	1
1.1. 什麼是重複資料刪除	2
1.2. 什麼是壓縮	4
1.3. 重複資料刪除與壓縮比較	7
1.4. 影響類型的因素	8
1.5. 資料縮減使用指南	9
2. 運作原理	11
2.1. 寫入資料	11
2.2. 資料縮減處理流程	12
2.3. 讀取資料	12
2.4. 覆蓋資料	13
3. 配置	14
3.1. 支援韌體版本	14
3.2. 授權	14
3.3. 限制	15

3.4.	配置資料縮減	16
4.	測試結果	17
4.1.	節省空間測試	17
4.2.	效能測試	18
5.	結論	20
6.	附錄	21
6.1.	適用於	21
6.2.	參考	21

圖表

圖表 1-1	資料縮減處理流程 (圖片來自紅帽部落格)	1
圖表 1-2	重複資料刪除 (圖片來自 GK_RAJ 貼文).....	2
圖表 1-3	重複資料刪除的處理流程 (圖片來自 ResearchGate 文章).....	3
圖表 1-4	資料壓縮類型	5
圖表 1-5	資料壓縮的處理流程 (圖來自 GITTA).....	6
圖表 1-6	重複資料刪除和壓縮的互補使用.....	7
圖表 1-7	每 TB 的 TCO (總體擁有成本) 與資料縮減率.....	10
圖表 2-1	資料縮減概述	11
圖表 2-2	資料縮減處理流程 (圖片來自紅帽部落格)	12

表格

表格 1-1	重複資料刪除與壓縮比較.....	7
表格 3-1	重複資料刪除和壓縮儲存池參數.....	15
表格 3-2	重複資料刪除和壓縮儲存池對應表.....	16
表格 4-1	節省空間測試結果.....	18
表格 4-2	效能測試結果	19

注意

文件中所包含資訊的準確性已被審查。但它可能包含印刷錯誤或技術不準確，這將定期對文件進行更改，而這些更改將納入該出版物的新版本。QSAN 可能會對產品進行改進或更改，所有功能和產品規格如有更改，恕不另行通知或承擔義務。本文件中的所有陳述、資訊和建議均不構成任何明示或暗示的擔保。

此處包含的任何效能資料都是在受控環境中確定的。因此，在其他作業環境中獲得的結果可能會有很大差異。在開發級系統上進行的一些測試，並無法保證這些測試在一般的系統上是相同的。此外，一些測量值可能是透過外推估計的。實際結果可能會有所不同。本文件的使用者應驗證其特定環境的適用資料。

此資訊包含日常商業作業中使用的資料和報告的範例。為了盡可能完整地說明它們，這些例子包含個人、公司、品牌和產品的名稱。所有這些名稱都是虛構的，與實際商業企業使用的名稱和地址如有任何相似之處，純屬巧合。

前言

技術支援

您是否有任何問題或需要幫助解決問題？請聯絡 QSAN 技術支援團隊，我們會盡快回覆給您。

- 透過網站：https://www.qsan.com/technical_support

- 透過電話：+886-2-77206355

(服務時間：09:30 - 18:00，週一至週五，UTC+8)

- 透過 Skype 通話，Skype ID：qsan.support

(服務時間：09:30 - 02:00，週一至週五，UTC+8，夏令時間：09:30 - 01:00)

- 透過電子郵件：support@qsan.com

資訊、提示和注意事項

本文件使用以下符號來提醒使用者注意重要的安全和操作資訊。



資訊

資訊提供有用的知識、定義或術語以供參考。



提示

提示為更有效地執行任務提供了有用的建議。



注意事項

注意事項表示不採取指定的措施可能會導致系統損壞。

1. 概述

這是一個資訊爆炸的時代，儘管擁有數 PB 的大量資料用於研究、人工智慧和分析聽起來很有好處，但過多的資料可能會導致儲存過載並阻礙高效能運算。然而，簡單地刪除資料並不是一個好的選擇，因為資料中的格式和順序對於分析至關重要。這就是資料縮減的用武之地，它轉換原始資料以減小其大小，同時保留其基本特徵。

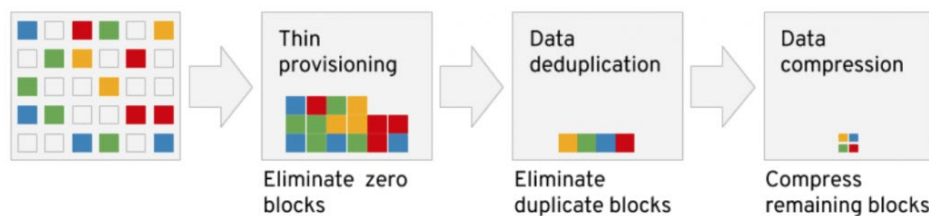
數據科學家使用多種技術來減少資料：

1. **重組**：透過刪除損壞或重複的項目、重新組織資料以消除不必要的物件以及識別非必要資訊來清理資料。
2. **編碼**：透過對特徵和模式進行編碼來建立更小的資料表示法。這可以是 "有損的" (透過損失一些保真度來減少檔案大小) 或 "無損的" (保持資料相同但檔案更小)。

資料縮減是一種透過消除冗餘資訊來減少儲存和傳輸的資料量的技術。此處理流程目的是提高儲存效率、降低成本並提高資料處理和分析的效能。無論採用哪種方法，資料縮減都是一個複雜的處理流程，工程師和科學家必須在空間節省、處理節省和資料保真度之間進行權衡。

一些常見的資料縮減類型是：

1. **精簡配置**：根據需要動態分配儲存空間，而不是預先分配。儘管需要更大的計算量，但此方法可以減少硬碟碎片等低效率問題。
2. **重複資料刪除**：刪除重複數據，包括刪除完全一致或表示相同資訊的記錄。
3. **壓縮**：應用演算法來減少資料的儲存空間。這可以在將資料移入儲存或靜態資料時完成，以節省更多空間。

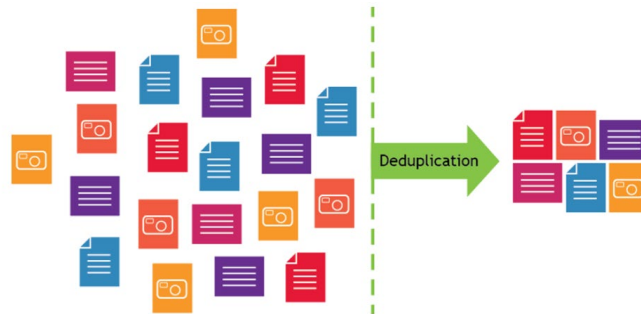


圖表 1-1 資料縮減處理流程 (圖片來自紅帽部落格)

透過利用這些技術，資料縮減可以顯著提高儲存效率和網路頻寬利用率。在本文件中，我們將深入描述重複資料刪除和壓縮。

1.1. 什麼是重複資料刪除

重複資料刪除是一種資料壓縮技術，用於消除資料的冗餘副本，從而減少所需的儲存空間量。此處理流程可確保僅儲存資料的唯一實體，而重複項將替換為原始資料的索引或指標。重複資料刪除可以在各個等級進行，例如檔案級或區塊級，具體取決於所分析的資料的粒度。



圖表 1-2 重複資料刪除 (圖片來自 GK_RAJ 貼文)

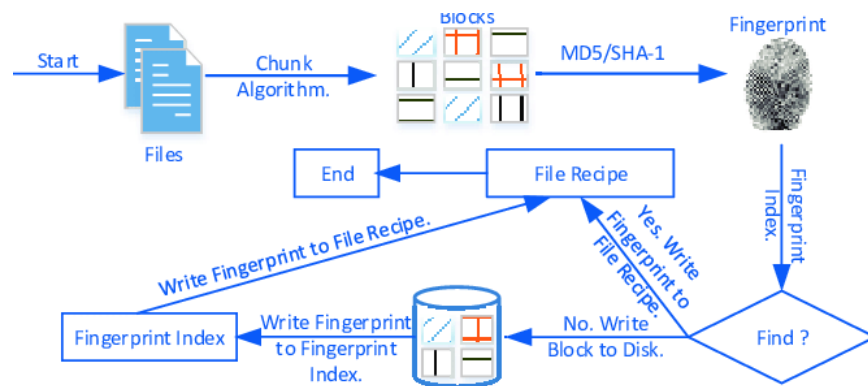
重複資料刪除有兩個等級：

1. **檔案級重複資料刪除**：比較整個檔案並刪除重複項目，在經常保存相同檔案的環境中很有用。
2. **區塊級重複資料刪除**：將檔案分成更小的區塊並刪除重複的區塊，比檔案層級的重複資料刪除更精細、更有效，特別是對於差異較小的大型資料集。

1.1.1. 重複資料刪除的處理流程

重複資料刪除的工作原理是識別和消除資料的冗餘副本，確保只儲存唯一的實體。下面簡要說明該處理流程的工作原理：

1. **資料分塊**：資料被分成更小的區塊或區段，這可以使用固定大小的區塊或可變大小的區塊來完成，具體取決於重複資料刪除方法。
2. **雜湊 (Hashing)**：每個區塊都透過雜湊演算法進行處理，產生唯一的識別碼，稱為雜湊值或指紋。該識別碼代表區塊的內容。
3. **比較**：將每個區塊的雜湊值與現有雜湊值的資料庫進行比較。如果在資料庫中找不到該雜湊值，則該區塊是唯一的，需要儲存。如果雜湊值已經存在，則表示該區塊是重複的。
4. **儲存**：新的區塊 (具有唯一的雜湊值) 儲存在儲存系統中。它們的雜湊值被新增到資料庫中以供將來比較。建立對已儲存唯一區塊的索引 (或指標)，而不是儲存重複的區塊。
5. **元資料 (Metadata) 管理**：維護元資料以追蹤原始資料的索引和結構。這允許系統在需要時重建原始資料。



圖表 1-3 重複資料刪除的處理流程 (圖片來自 ResearchGate 文章)

重複資料刪除可以透過兩種主要的方式實現：線內重複資料刪除和後處理重複資料刪除。每種方法都有自己的優點和使用案例，這取決於儲存環境的特定要求和效能需求。

1. **線內重複資料刪除**：它發生在資料寫入處理流程中。當資料寫入儲存系統時，系統會進行即時分析，並在儲存之前識別並消除重複資料。
2. **後處理重複資料刪除**：它發生在資料寫入儲存之後。資料首先以其原始形式寫入，然後在一個單獨的處理程式掃描資料、識別重複項目並消除。

線內重複資料刪除和後處理重複資料刪除都為減少儲存空間提供了寶貴的優勢，但它們滿足不同的需求。線內重複資料刪除可立即節省空間，並且在高冗餘環境中有益，儘管它可能會引入一些

寫入延遲。後處理重複資料刪除可以避免影響寫入效能，但在初始寫入後才實現節省空間，使其適合歸檔和寫入效能至關重要的系統。

1.1.2. 重複資料刪除的優點和挑戰

以下列出使用重複資料刪除的好處：

- **儲存效率：**透過消除冗餘數據，大幅減少所需的儲存空間。
- **節省成本：**透過最大限度地利用現有儲存資源來降低儲存基礎設施成本。
- **提升備份和復原：**透過減少需要傳輸和儲存的資料量來加快備份和復原程序。
- **增強的資料管理：**透過減少要處理的資料量來簡化資料管理。

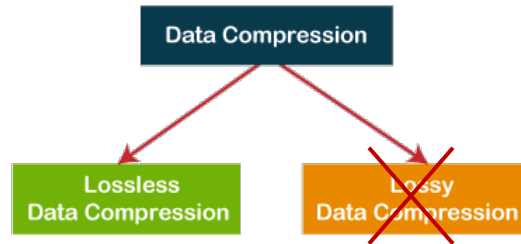
以下是使用重複資料刪除的挑戰：

- **處理負擔：**初始重複資料刪除可能需要大量的處理能力和時間。
- **系統效能：**如果管理不當，連續的重複資料刪除程序可能會影響系統效能。
- **複雜性：**實施和維護重複資料刪除系統可能很複雜，並且可能需要專門的工具和專業知識。

重複資料刪除是一項可透過消除冗餘資料副本來優化儲存效率並降低成本的強大技術。透過僅儲存唯一實體加以參考取代重複項目，重複資料刪除可大幅減少儲存佔用空間並增強資料管理。儘管存在實施方面的挑戰以及對系統效能的潛在影響，重複資料刪除的優勢使其成為現代資料儲存策略的重要組成部分。

1.2. 什麼是壓縮

儲存中的壓縮是一種透過更有效地編碼來減少資料大小的技術。此處理流程有助於節省儲存空間、降低成本並提高資料傳輸速度。壓縮可應用於各種類型的資料，包括文字、圖像、音訊和影片，並且常用於檔案儲存和資料傳輸。



圖表 1-4 資料壓縮類型

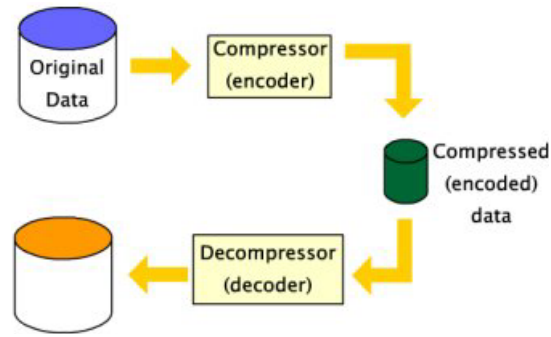
資料壓縮有兩種類型：

1. **無損壓縮**：減少資料大小而不遺失任何資訊，可以從壓縮資料完美地重建原始資料。它用於文字文件、可執行檔案以及完整性和精確複製都至關重要的任何資料。例如：ZIP 檔案、PNG 影像和 FLAC 音訊檔案。
2. **有損壓縮**：透過刪除一些資訊來減少資料大小，這可能會導致品質損失，且原始資料無法完美重建。它用於可以接受品質降低以節省大量空間的多媒體文件，例如照片、音樂和影片。例如：JPEG 影像、MP3 音訊檔案和 MPEG 影片檔案。有損壓縮不能用於儲存，因為檔案無法恢復到原始形式。

1.2.1. 資料壓縮的處理流程

資料壓縮的運作原理是對資料進行編碼和解碼，確保無損資料壓縮不會損失任何資料品質，並且解壓縮的檔案是原始檔案的精確副本。以下是該處理流程如何運作的簡化解釋：

1. **資料分析**：分析資料以識別模式、重複和冗餘。
2. **編碼**：應用演算法以較短的表示法取代重複的元素。常見的演算法包括霍夫曼編碼、LZW (Lempel-Ziv-Welch) 和 RLE (Run-Length Encoding 運行長度編碼)。
3. **儲存或傳輸**：將壓縮資料與字典或金鑰一起儲存或傳輸以進行解壓縮。



圖表 1-5 資料壓縮的處理流程 (圖來自 GITTA)

同樣地，資料壓縮可以線內和後處理來實現。線內壓縮和後處理壓縮都為減少儲存空間提供了寶貴的好處，但它們滿足不同的需求和場景。線內壓縮可以立即節省空間並提高寫入效能，但它可能會帶有延遲並且需要複雜的實作。後處理壓縮可避免初始寫入延遲並允許靈活調度，但空間節省是稍後實現的，並且需要額外的臨時儲存容量。

1.2.2. 壓縮的好處和挑戰

下面列出了使用壓縮的一些好處：

- **儲存效率：**減少資料所需的儲存空間。
- **節省成本：**降低儲存基礎設施成本並減少資料傳輸費用。
- **更快的資料傳輸：**壓縮資料通過網路傳輸所需的時間更少，從而提高了效能。
- **提升備份和復原：**較小的備份檔案可加快備份和復原程序。

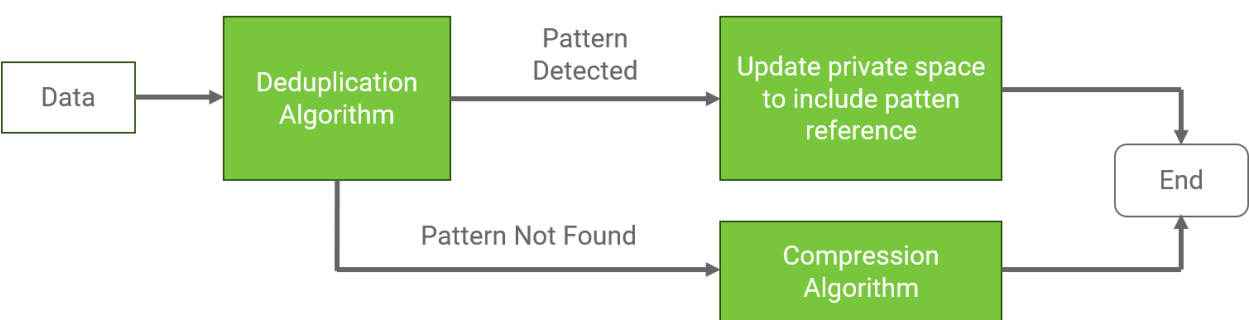
以下是使用壓縮的挑戰：

- **處理負擔：**壓縮和解壓縮資料需要額外的處理能力和時間。
- **複雜性：**實施和管理壓縮演算法可能很複雜，並且可能需要專業知識。

資料壓縮是優化儲存和傳輸效率的強大技術。透過更有效地編碼資料，壓縮減少了儲存空間需求並提高了數據傳輸速度。

1.3. 重複資料刪除與壓縮比較

在重複資料刪除中，資料根據其中的公共區塊進行聚集。每個區塊的單一版本被保留，而其他出現的版本被散列或使用指標索引。另一方面，在壓縮中，附加資料、空格等被消除以減少資料檔案的大小。因此，重複資料刪除的重點是去除重複；壓縮的重點是少資料的大小。重複資料刪除對於備份和歸檔有效；壓縮對於檔案傳輸和儲存有效。



圖表 1-6 重複資料刪除和壓縮的互補使用

對於大小減少的比率，壓縮聲稱將資料大小減少到 2.5:1 的比率，正如一些基於可用資料檔案類型的程式所聲稱的那樣。然而，透過重複資料刪除，資料將會發生顯著變化。減速比範圍為 4:1 至 10:1，對於某些材料類型甚至更高。

表格 1-1 重複資料刪除與壓縮比較

	重複資料刪除	壓縮
定義	消除資料的冗餘副本	透過更有效地編碼來減少資料大小
方法	識別並儲存唯一的資料塊，用索引替換重複項	對資料進行編碼以使用更少的位元，使用演算法更緊湊地表示資料

效率	對於具有大量重複資料的環境非常有效	對大多數類型的資料有效，但效率不同
節省空間	根據重複資料量，縮減比例範圍為 4:1 到 10:1	一致的空間節省，根據資料類型和演算法可能會較為顯著，將資料大小縮減到 2.5:1 的比率
處理負擔	可能很高，特別是對於線內重複資料刪除	各不相同；即時 (線內) 壓縮會造成延遲
使用案例	• 備份儲存 • 雲端儲存 • VDI (虛擬桌面基礎架構)	• 檔案儲存 • 資料傳輸 • 多媒體檔案
互補使用	可結合壓縮進一步優化儲存效率	可與重複資料刪除結合使用，以獲得額外的儲存縮減優勢

重複資料刪除和壓縮都有各自的優點和限制，為了互補使用，兩者可以結合使用以獲得最大益處。

1.4. 影響類型的因素

在為資料縮減方法設定容量節省目標時，必須考慮以下影響重複資料刪除和壓縮效率比的決定性因素，這對儲存效能至關重要：

- **資料類型：**它在決定重複資料刪除或壓縮的程度方面起著關鍵作用。快照 (例如在 VDI 環境中) 預計將提供更高的容量最佳化率。對於資料庫來說，在應用層面已經有一定程度的冗餘消除。因此，資料庫可能不會透過重複資料刪除和壓縮產生顯著的節省。

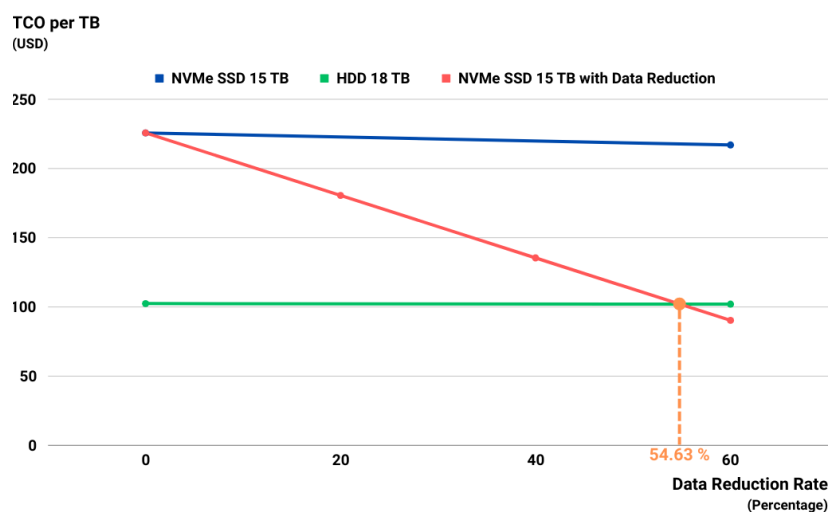
- **資料更改率：**資料變更越少，執行重複資料刪除和壓縮就越容易、越快。對資料進行的每次變更並儲存到硬碟都需要單獨的檢查週期來識別重複的資料集並執行壓縮演算法。這隨後會增加伺服器負擔。
- **資料備份頻率：**尤其是完整備份的情況下，備份頻率越高，產生的冗餘資料就越多。這將在重複資料刪除和壓縮過程中節省更多空間。

資料縮減的效率可以用原始資料大小與縮減資料大小的比率來衡量。例如，當重刪比為 10:1 時，100 GB 的原始資料只需要 10 GB 的儲存容量，從而節省 90% 的空間。此比率越高，節省的容量就越大。這取決於軟體執行重複資料刪除和壓縮的能力，以及上述因素。

1.5. 資料縮減使用指南

儘管資料縮減的應用範圍有限，但在大多數情況下仍更具成本效益。根據 SNIA 報告，"傳統硬碟：NVMe SSD" 每 TB 的 TCO (總體擁有成本) 百分比約為 45.37%。當應用資料縮減率達到 54.63% 或更高時，重複資料刪除策略將獲得比部署新儲存裝置更好的 TCO，如下圖所示。由於應用程式更適合重複資料刪除策略，因此可以實現更好的 TCO。此外，資料縮減策略可能會導致效能下降，但在儲存環境中部署 SSD 的體驗仍然會變得更加有力。

TCO per TB v.s. Data Reduction Rate



圖表 1-7 每 TB 的 TCO (總體擁有成本) 與資料縮減率

2. 運作原理

資料縮減功能結合了各種空間效率演算法，以最大限度地減少資料集所需的儲存空間。此功能包括重複資料刪除和壓縮演算法。資料縮減對於區塊和檔案儲存卷的作用相同。它使用軟體演算法來分析並實現儲存資源內的空間節省，並在記憶體快取和硬碟之間線內發生。本章介紹了資料縮減的運作原理，分解幾項運作，並分別進行描述。



圖表 2-1 資料縮減概述

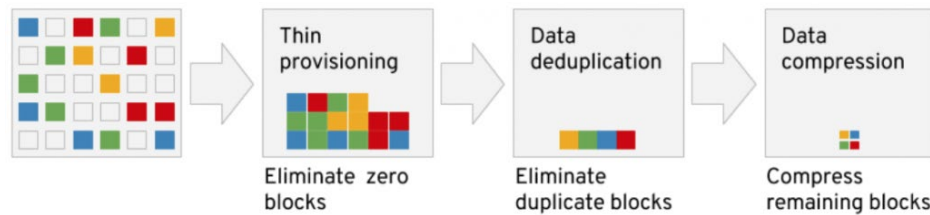
2.1. 寫入資料

當資料寫入儲存時，首先將其保存在記憶體快取中。在將寫入內容儲存到記憶體快取之前，儲存會檢查目標儲存資源中是否有足夠的空間，並為 I/O 分配空間。為了確保最快的回應，寫入 I/O 的資料縮減演算法在此階段不執行，而是立即向主機確認寫入。

對於支援資料縮減的儲存，該處理流程發生在主動記憶體快取清理期間或記憶體將快取頁面排清到硬碟時。在此處理流程中，多個檔案或區塊透過演算法被組合和處理。根據演算法的結果，分配儲存卷內的空間，並在必要時將資料寫入硬碟。

2.2. 資料縮減處理流程

下圖概述了在儲存卷中啟用重複資料刪除和壓縮的資料縮減功能。它結合了三種技術 — 零區塊消除、重複資料刪除和資料壓縮 — 來減少資料佔用。



圖表 2-2 資料縮減處理流程 (圖片來自紅帽部落格)

這些是資料縮減處理流程中的關鍵階段，使引擎能夠減少儲存上的資料佔用空間。它在線內且動態地套用這些階段。以下討論資料縮減功能中的每種演算法。

1. **零區塊消除**：在初始階段，任何完全由零組成的區塊將被識別並僅記錄在元資料中。在此階段，該處理流程僅允許包含非全零的資料被過濾到下一個處理階段。
2. **重複資料刪除**：在第二階段，處理傳入的資料以確定其是否冗餘 (先前已寫入的資料)。透過作為資料縮減引擎的一部分提供的重複資料刪除服務核心模組所維護的元資料來檢查該資料的冗餘性。任何發現冗餘的資料塊都不會被寫出。相反地，元資料將被更新以指向已儲存在媒體上的區塊的原始副本。
3. **壓縮**：初始零消除和重複資料刪除階段完成後，將對個別的資料套用壓縮。然後，壓縮資料被打包成固定長度的區塊並儲存在媒體上。這也提高了從儲存讀取資料的效能，因為單一實體區塊可以包含許多壓縮區塊。

2.3. 讀取資料

當從已啟用資料縮減的儲存資源請求讀取運作時，系統首先確定資料所在的位置。它檢查資料是否位於記憶體快取中或硬碟上，以及資料是否處於原始大小或已進行重複資料刪除或壓縮。如果

資料以其原始形式存在於記憶體快取中，則直接傳送到主機。如果資料在硬碟上保持原始形式，則執行標準讀取運作，將資料複製到記憶體快取中，然後將其傳送到主機。

如果資料的大小已減小，則必須在記憶體快取中將其還原為原始形式，然後再傳送到主機。對於已進行重複資料刪除的資料，如果它們包含共同模式，則會重新建立它們。然後將資料傳送到主機。如果資料已壓縮，則在傳送到主機之前必須解壓縮。如果壓縮資料已經在記憶體快取中，則將其解壓縮到臨時位置，發送到主機，並清除臨時位置。如果壓縮資料位於硬碟上，則會將其讀入記憶體快取，解壓縮到臨時位置，然後傳送至主機。資料永遠不會在硬碟本身上解壓縮，因為這會抵消已節省的儲存空間。

2.4. 覆蓋資料

當收到先前寫入的區塊或檔案的更新時，系統會檢查它是否節省了空間以及更新是否可以實現進一步的節省。資料經過資料縮減處理流程以確定任何潛在的空間節省。如果新區塊或檔案與已知模式相符，則資源內的私有空間將使用新模式資訊進行更新。如果舊的有壓縮節省或以其原始形式存儲，則這些空間將被釋放以供重複使用。如果無法進行重複資料刪除，則資料將透過壓縮邏輯發送。

如果壓縮減少了資料大小，系統將決定把資料儲存在硬碟中的位置。如果新資料大小比以前大，則分配新的空間。如果大小未更改或較小，則可以將資料寫入現有分配的區塊。這種方法可以防止碎片，提高效能並節省空間。如果分配了新區塊，則釋放舊區塊以供重複使用。如果無法進行重複資料刪除或壓縮節省，則資料將以其原始大小寫入，可能會覆寫原始資料。

3. 配置

本章將討論如何設定資料縮減。

3.1. 支援韌體版本

以下總結支援資料縮減的韌體版本和限制。

韌體版本

- XEVO
 - 韌體版本 3.0.0 及更高版本，需有授權
- QSM
 - 韌體版本 4.1.4 及更高版本，需有授權

3.2. 授權

必須購買重複資料刪除和壓縮授權才能啟用該功能，另外也有 30 天的免費評估試用授權。

3.2.1. 啟用重複資料刪除和壓縮授權

重複資料刪除和壓縮功能是選配的。使用前，您必須啟用重複資料刪除和壓縮授權。選擇**系統選**項和**維護**子選項，然後點擊**授權**窗格以允許使用者啟動授權。

點擊**請求授權**按鈕下載檔案並發送給您當地的銷售人員以獲得授權密鑰。取得授權密鑰後，點擊圖示選擇其密鑰，然後點擊**套用**按鈕以啟用。每個許可證密鑰都是唯一的且專用於特定系統。如果授權處於活躍狀態，狀態將顯示為已**啟用**。啟用授權後，必須手動重新啟動系統才能生效。

3.3. 限制

關於資料縮減功能有一些注意事項。

- 重複資料刪除和壓縮需要授權。
- 僅支援全快閃儲存池，甚至包括備用硬碟。
- 重複資料刪除和壓縮將同時啟用或停用。
- 建立儲存池時啟用重複資料刪除和壓縮，然後在建立儲存卷時啟用它們。
- 支持硬碟漫遊。
- 試用授權到期後，所有具有重複資料刪除和壓縮功能的儲存卷將會停用。

儲存池的總容量將根據系統記憶體大小而受到限制。以下是重複資料刪除和壓縮儲存池的參數。

表格 3-1 重複資料刪除和壓縮儲存池參數

項目	值
每個系統的最大儲存池容量	系統記憶體 x 1,024
每個系統的最大儲存池數	2
每個系統的最大儲存卷數	系統記憶體 / 4

以下是重複資料刪除和壓縮儲存池記憶體、容量、與儲存卷數的對應表。

表格 3-2 重複資料刪除和壓縮儲存池對應表

每個控制器的系統記憶體	每個系統的最大重複資料刪除 和壓縮儲存池容量	每個系統的最大儲存卷 數
8 GB	8 TB	2
16 GB	16 TB	4
32 GB	32 TB	8
64 GB	64 TB	16
128 GB	128 TB	32
256 GB	256 TB	64
512 GB	512 TB	128
1,024 GB	1,024 TB	256

3.4. 配置資料縮減

有關重複資料刪除和壓縮池的操作，請參閱 QSAN 網站上的 [XEVO 3 或 QSM 4 軟體手冊](#)。

4. 測試結果

本章將介紹資料縮減的測試結果。

4.1. 節省空間測試

此測試展示透過資料縮減實現的空間節省。

測試設備及配置

- 伺服器
 - 型號：華碩 RS700-E6/ERS4 (中央處理器：Intel Xeon E5620 2.4 GHz / 記憶體：24 GB)
10 GbE 主機匯流排介面卡：Broadcom BCM57810 NetXtreme II 10 GbE
作業系統：Windows Server 2019
- 儲存設備
 - 型號：QSAN XN8126D
記憶體：每個控制器 16 GB
韌體版本：4.1.0
固態硬碟：7 顆 KIOXIA PX05SMB 系列，PX05SMB080，800 GB，SAS 12 Gb/s
 - 全快閃儲存池：1 個 RAID 6 儲存池，控制器 1 中有 7 顆固態硬碟
儲存卷：1 個 500 GB，啟用重複資料刪除和壓縮，另一個
- 測試檔案
 - 大影片檔案：20 GB
 - 零碎小檔案：超過 11 萬個檔案，每個檔案大小在 1 KB 到 218 KB 之間，總計 6.69 GB
 - VMware VMDK 檔案：總計 21.4 GB

測試腳本及結果

1. 建立一個啟用重複資料刪除和壓縮的儲存池，然後建立一個也啟用了該功能的儲存卷。
2. 將大影片檔案寫入儲存池，並將佔用的容量記錄為第 1 次寫入。
3. 將同一個大影片檔案寫入不同的目錄，然後將佔用容量記錄為第 2 次寫入。
4. 對於零碎小檔案和 VMware VMDK 檔案重複步驟 2 和 3。

表格 4-1 節省空間測試結果

測試項目	第 1 次寫入	第 2 次寫入	空間節省率	縮減比
		原始 / 實際		
大影片檔案	20 GB	40 GB / 36 GB	10%	1.1 : 1
零碎小檔案	6.69 GB	13.38 GB / 7 GB	49%	5.1 : 1
VMware VMDK 檔案	21.4 GB	42.8 GB / 29.6 GB	86%	6.9 : 1

4.2. 效能測試

此測試展示資料縮減對效能的影響。

測試設備及配置

- 伺服器
 - 型號：華碩 RS700-E6/ERS4 (中央處理器：Intel Xeon E5620 2.4 GHz / 記憶體：24 GB)
 - 10 GbE 主機匯流排介面卡：Broadcom BCM57810 NetXtreme II 10 GbE
 - 作業系統：Windows Server 2019
- 儲存設備

- 型號：QSAN XN8126D
記憶體：每個控制器 16 GB
韌體版本：4.1.0
固態硬碟：7 顆 KIOXIA PX05SMB 系列，PX05SMB080，800 GB，SAS 12 Gb/s
- 全快閃儲存池：1 個 RAID 6 儲存池，控制器 1 中有 7 顆固態硬碟
儲存卷：1 個 500 GB，啟用重複資料刪除和壓縮，另 1 個 500 GB 停用
- 測試檔案
 - 大影片檔案：20 GB
 - 零碎小檔案：超過 11 萬個檔案，每個檔案大小在 1 KB 到 218 KB 之間，總計 6.69 GB

測試腳本及結果

1. 建立一個啟用重複資料刪除和壓縮的儲存池，然後建立 2 個儲存卷。一個儲存卷啟用功能，而另一個儲存卷停用功能。
2. 將大視訊檔案複製到這 2 個儲存卷中並記錄傳輸速度。
3. 對於碎片小檔案重複步驟 2。

表格 4-2 效能測試結果

測試項目	啟用	停用	效能影響	節省時間
大影片檔案	350 MB/s	750 ~ 800 MB/s	-53% ~ -58%	不適用
零碎小檔案	700 ~ 800 KB/s (需 2 小時)	1 ~ 1.5 MB/s (需 1.5 小時)	-20% ~ -53%	25%

5. 結論

資料縮減是資料管理的重要策略，可幫助企業更有效地處理不斷增加的資料量。透過實施重複資料刪除和壓縮等技術，組織可以顯著節省成本、提高效能並優化其儲存基礎架構。這整體的方法可確保資料儲存在面對不斷增長的資料需求時保持高效且可管理。

6. 附錄

6.1. 適用於

重複資料刪除和壓縮需要**授權**才能啟用該功能。我們還提供 30 天免費**試用授權**以供評估。重複資料刪除和壓縮功能適用於下列機型。

- XEVO
 - 韌體版本 3.0.0 及更高版本，需有**授權**
- QSM
 - 韌體版本 4.1.4 及更高版本，需有**授權**

6.2. 參考

文件

- [XEVO 3 軟體手冊](#)
- [QSM 4 軟體手冊](#)