XCubeSAN系列技術白皮書

RAIDEE技術

QSAN廣盛科技
www.QSAN.com

版權

©版權所有 2019 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。

2019 年 9 月

此版本適用於 QSAN XCubeSAN系列。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、XCubeSAN和 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。

注意

此 XCubeSAN系列技術白皮書適用於以下 XCubeSAN型號：

XCubeSAN儲存系統 4U 19U機架式機箱型號

型號名稱	控制器類型	外形、槽位計數和機架單元
XS5224D	雙控制器	LFF 24磁碟 4U 機箱
XS3224D	雙控制器	LFF 24磁碟 4U 機箱
XS3224S	單控制器	LFF 24磁碟 4U 機箱
XS1224D	雙控制器	LFF 24磁碟 4U 機箱
XS1224S	單控制器	LFF 24磁碟 4U 機箱

XCubeSAN儲存系統 3U 19U機架式機箱型號

型號名稱	控制器類型	外形、槽位計數和機架單元
XS5216D	雙控制器	LFF16 磁碟 3U 機箱
XS3216D	雙控制器	LFF16 磁碟 3U 機箱
XS3216S	單控制器	LFF16 磁碟 3U 機箱
XS1216D	雙控制器	LFF16 磁碟 3U 機箱
XS1216S	單控制器	LFF16 磁碟 3U 機箱

XCubeSAN儲存系統 2U 19U機架式機箱型號

型號名稱	控制器類型	外形、槽位計數和機架單元
XS5212D	雙控制器	LFF12 磁碟 2U 機箱
XS5212S	單控制器	LFF 12磁碟 2U 機箱
XS3212D	雙控制器	LFF 12磁碟 2U 機箱

XS3212S	單控制器	LFF 12磁碟 2U 機箱
XS1212D	雙控制器	LFF 12磁碟 2U 機箱
XS1212S	單控制器	LFF 12磁碟 2U 機箱
XS5226D	雙控制器	SFF 26 磁碟 2U 機箱
XS5226S	單控制器	SFF 26磁碟 2U 機箱
XS3226D	雙控制器	SFF 26磁碟 2U 機箱
XS3226S	單控制器	SFF 26磁碟 2U 機箱
XS1226D	雙控制器	SFF 26磁碟 2U 機箱
XS1226S	單控制器	SFF 26磁碟 2U 機箱

文件中所包含資訊的準確性已被審查。但它可能包括印刷錯誤或技術不[redacted]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，可以通過外推估計一些測量值，實際結果可能有所不同，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，這些例子包括個人、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完全是巧合。

 ©版權所有 2019 QSAN廣盛科技保留所有權利。

目錄

RAID EE技術

執行摘要

已經存在超過 30 年的 RAID架構正在經歷一波轉型。對於 TB 級大容量硬碟，原始 RAID技術無法解決重建時間過長的問題。基於傳統區塊技術的新一代 RAID技術，我們稱之為 RAID EE，這是被視為解決傳統 RAID缺陷的途徑。

資訊：

RAID EE技術在 SANOS韌體版本 1.3.0 中可用，並且在 SANOS韌體版本 1.4.1 中效能有大幅度　提升。

讀者

本文件適用於有興趣了解 RAID EE以解決重建時間過長問題的 QSAN客戶和合作夥伴。我們假定讀者熟悉 QSAN產品並具有一般 IT 經驗，包括作為系統或網路管理員的知識。如有任何疑問，請參閱產品的使用手冊，或聯絡 QSAN技術支援以獲得進一步的幫助。

概觀

RAID（獨立磁碟冗餘陣列）將基於特定演算法組合多個獨立的物理磁碟，以形成虛擬邏輯磁碟，從而提供更大容量、更高效能或更好的資料容錯能力。RAID作為一種成熟可靠的資料保護標準，已成為儲存系統的基礎技術。然而，隨著近年來磁碟對資料儲存的需求快速增長以及高效能應用的出現，傳統 RAID逐漸暴露出其缺陷。

隨著硬碟容量的增加，重建 RAID資料所需的時間也大大增加。這是當今企業儲存管理中最麻煩的問題之一。在過去幾天硬碟容量只有 10GB 到 100GB 的情況下，RAID重建的工作可以在 10 分鐘甚至 10 分鐘以上完成，這在沒有特別關注的情況下還不是問題。 但是，隨著磁碟容量增長到數百 GB 甚至 TB，RAID重建時間增加到數小時甚至數天，這成為儲存管理中的主要問題。

例如，一個傳統 RAID 5內含有 8 加 1 顆同位檢查的 6TB NL-SAS磁碟上需要 2.5 天才能重建資料。重建過程會消耗系統資源，從而降低應用程式系統的整體效能。如果使用者限制重建優先權，則重建時間將更長。重要的是，在耗時的重建過程中，大量的存取操作可能導致儲存池中其他磁碟的故障，從而大大增加了磁碟故障的可能性和資料丟失的風險。

傳統 RAID 架構的局限性

傳統的 RAID架構由一定數量的磁碟組成磁碟組 (也稱為 RAID組)。您還可以將某些磁碟指定為空閒的熱備援磁碟。儲存池被分組以提供儲存卷的容量，然後最終將 LUN 映射到主機以成為主機上的儲存空間。

這種 RAID架構有幾個局限性：

- [illegible]加載的資料集中在備援磁碟上形成瓶頸。
- 其次，儲存卷資料存取僅限於屬於磁碟組的成員磁碟；這限制了主機的效能，因為儲存裝置正在執行存取和重建 I/O。

為什麼 RAID 重建耗費時間

隨著磁碟容量的增長，RAID重建時間呈線性增長，當使用具有 4TB 以上硬碟容量的 RAID磁碟時，將傳統 RAID架構所需的重建時間提高到數十小時。

有幾個因素會影響 RAID重建時間：

- **硬碟容量：**硬碟容量構成磁碟組，硬碟容量越大，重建時間越長。
- **磁碟數量：**[illegible]援磁碟所需的時間。磁碟越多，重建時間越長。
- **重建作業優先權：**在 RAID重建期間，系統仍需要假設對前端主機的 I/O 存取。分配給 RAID重建作業的優先權越高，重建速度越快，但前端主機獲得 I/O 效能的能力越低。
- **快速重建：**啟用快速重建功能只需要重建儲存卷的實際容量，未使用的磁碟組空間不用重建。如果儲存卷僅使用磁碟組中的部分空間，則將縮短重建時間。
- RAID **等級：**具有直接區塊到區塊複製的 RAID 1和 RAID 10將比具有同位檢查 RAID 5和 RAID 6重建地更快。

鑑於每個磁碟可能出現故障，磁碟組中包含的磁碟越多，累積故障的可能性就越大，因此磁碟組中的磁碟數量有上限的限制。與以前的因素相比，磁碟容量對重建速度的影響越來越大，這已成為主要因素。如此長的重建時間顯然是任何使用者都不能接受的。為了解決傳統 RAID 的問題，我們實現了 RAID EE技術。

 ©版權所有 2019 QSAN廣盛科技保留所有權利。

運作原理

RAID EE在磁碟組中新增了更多備援磁碟，我們將其稱為 RAID EE**備援磁碟**，以分離原始的全域、本地和專用備援磁碟。備援區域分佈在磁碟組的每個條帶中，並透過磁碟旋轉分佈在磁碟組中。磁碟組中的磁碟發生故障時，缺少的資料將重建到保留的備援區中。由於集合中的所有磁碟都是重建資料的目標，因此傳統 RAID重建的瓶頸消失了，重建效能得到了顯著提升。如果新增了新磁碟，則備援區中的資料將複製回新加入的磁碟。

為 RAID EE提供了四個新的 RAID等級，有：

- RAID 5EE(E 代表增強型)，至少需要 4 顆磁碟和一顆 RAID EE備援磁碟，可以容忍 2 顆磁碟故障。新增更多 RAID EE備援磁碟將容忍更多磁碟故障。
- RAID 6EE至少需要 5 顆磁碟。
- RAID 50EE至少需要 7 顆磁碟。
- RAID 60EE至少需要 9 顆磁碟。

資訊：

磁碟組中的 RAID EE備援磁碟數量為 1 到 8 顆磁碟。

帶有 1 顆 RAID EE備援磁碟的 RAID 5EE範例

舉一個例子來描述它是如何工作的。以下範例是具有 5 顆磁碟的 RAID 5EE，4 顆磁碟用於 RAID磁碟，另外一顆磁碟用於 RAID EE備援磁碟。初始化後，資料區塊分配如下。P 代表同位檢查，S 代表 RAID EE備援磁碟，現在它的內容是空的。

圖表 1　　RAID 5EE的資料區塊分佈

假設磁碟 2 發生故障。RAID 5EE處於降級模式。

D1	D2	D3	D4	D5
1	2	3	P	S
S	4	5	6	P
P	S	7	8	9
10	P	S	11	12
13	14	P	S	15

圖表 2　　磁碟 2 已故障

故障磁碟中的資料將被重建回備援區域。此操作稱為 EE重建。重建後，分佈的資料就像 RAID 5，它可以容忍另一個故障的磁碟。我們可以想像，RAID EE備援磁碟越多，重建速度就越快。

©版權所有 2019 QSAN廣盛科技保留所有權利。

D1		D3	D4	D5
1		3	P	2
4		5	6	P
P		7	8	9
10		P	11	12
13		P	14	15

圖表 3　　故障磁碟中的資料重建回空的區塊

當新磁碟加入 RAID EE磁碟組時，備援區中重建的資料[redacted]磁碟。此操作稱為**回拷**。

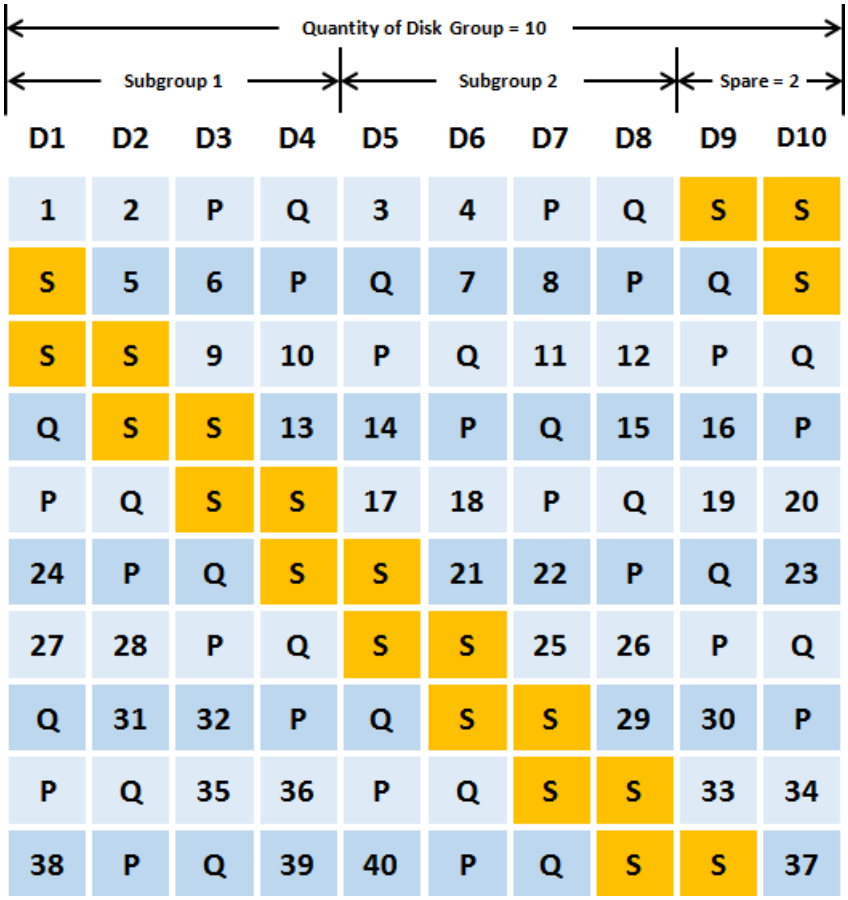
D1	D2	D3	D4	D5
1	2	3	P	S
S	4	5	6	P
P	S	7	8	9
10	P	S	11	12
13	14	P	S	15

圖表 4　　資料被回拷

回拷後，它回到 RAID 5EE的正常狀態。

帶有 2 顆 RAID EE備援磁碟的 RAID 60EE範例

再舉一個帶有 10 顆磁碟的 RAID 60EE範例。8 顆磁碟用於 RAID磁碟，2 顆磁碟用於 RAID EE備援磁碟。初始化後，資料區塊分配如下。



圖表 5　　RAID 60EE的資料區塊分佈

RAID 60EE的重建和回拷與上述類似；這裡不再重複。

RAID EE等級摘要

以下是 RAID EE等級摘要。

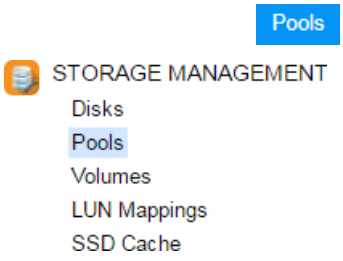
©版權所有 2019 QSAN廣盛科技保留所有權利。

表格 1　　RAIDEE等級摘要

	RAID 5EE	RAID 6EE	RAID 50EE	RAID 60EE
最少磁碟數量	4	5	7	9
容錯 (G =子組，S = RAID EE 備援磁碟 = 1 ~ 8)	2 ~ 9 顆磁碟故障(			

配置 RAID EE儲存池

RAID EE儲存池的操作。



圖表 6　　儲存池功能子菜單

建立 RAID EE儲存池

以下是建立配置 4 顆磁碟的 RAID 5EE儲存池範例。在首次建立儲存池時，包含一個磁碟組，且磁碟組中的最大磁碟數為 64。

1. 選擇**儲存池**功能子菜單，點擊**建立儲存池**按鈕。首先掃描可用磁碟。

提示：

如果您的系統有超過 200 顆磁碟，掃描磁碟可能需要 20 ~ 30 秒。請耐心等待。

©版權所有 2019 QSAN廣盛科技保留所有權利。

圖表 7　　建立 RAID EE儲存池步驟 1

2. 選擇**儲存池類型**
3. **儲存**　　15　　元為 [A~Z | a~z | 0~9 | -_<>]。
4. 從下拉列表中選擇**首選控制器**　此　　池中的後端 I/O 資源將由您指定的首選控制器處理，此選項在安裝雙控制器時可以使用。
5. 點選**啟用** SED**儲存池**複選框　啟用 SED　　池將使用安全的 SED建立儲存池，不支援混合 SED和非 SED磁碟在同一個儲存池中。
6. 點擊**下一步**