



# 高可用性架構

## ONTAP Select

NetApp  
January 29, 2026

# 目錄

高可用性架構 . . . . .	1
ONTAP Select 高可用性配置 . . . . .	1
雙節點 HA 與多節點 HA . . . . .	2
雙節點 HA 與雙節點延伸 HA (MetroCluster SDS) . . . . .	3
ONTAP Select HA RSM 和鏡像聚合 . . . . .	3
同步複製 . . . . .	3
鏡像聚合 . . . . .	3
寫入路徑 . . . . .	4
ONTAP Select HA 增強資料保護 . . . . .	6
磁碟心跳 . . . . .	6
HA 郵箱發帖 . . . . .	6
HA 心跳 . . . . .	6
HA 故障轉移與恢復 . . . . .	7

# 高可用性架構

## ONTAP Select高可用性配置

探索高可用性選項，為您的環境選擇最佳的 HA 配置。

儘管客戶開始將應用程式工作負載從企業級儲存設備遷移到運行在商用硬體上的基於軟體的解決方案，但對彈性和容錯能力的期望和需求並未改變。提供零復原點目標 (RPO) 的高可用性 (HA) 解決方案可保護客戶免受基礎架構堆疊中任何元件故障導致的資料遺失。

很大部分 SDS 市場建立在無共享儲存的概念之上，軟體複製透過在不同的儲存孤島中儲存多個用戶資料副本來提供資料彈性。ONTAP ONTAP Select基於此前提構建，使用ONTAP提供的同步複製功能 (RAID SyncMirror ) 在叢集內儲存額外的使用者資料副本。這發生在 HA 對的環境中。每個 HA 對都儲存兩個使用者資料副本：一個在本地節點提供的儲存上，另一個在 HA 合作夥伴提供的儲存上。在ONTAP Select叢集中，HA 和同步複製綁定在一起，兩者的功能不能分離或獨立使用。因此，同步複製功能僅在多節點產品中可用。

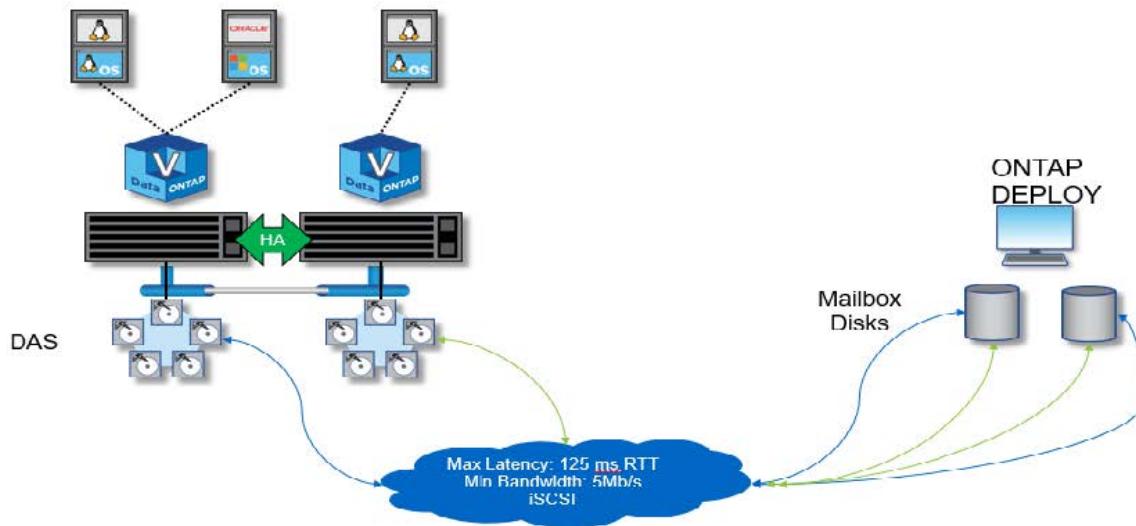


在ONTAP Select叢集中，同步複製功能是 HA 實現的功能，而不是非同步SnapMirror 或SnapVault複製引擎的替代品。同步複製不能獨立於 HA 使用。

ONTAP Select HA 部署模型有兩種：多節點叢集（四節點、六節點或八節點）和雙節點叢集。雙節點ONTAP Select叢集的顯著特徵是使用外部調解服務來解決腦裂問題。ONTAP Deploy ONTAP 機器可作為其配置的所有雙節點 HA 對的預設調解器。

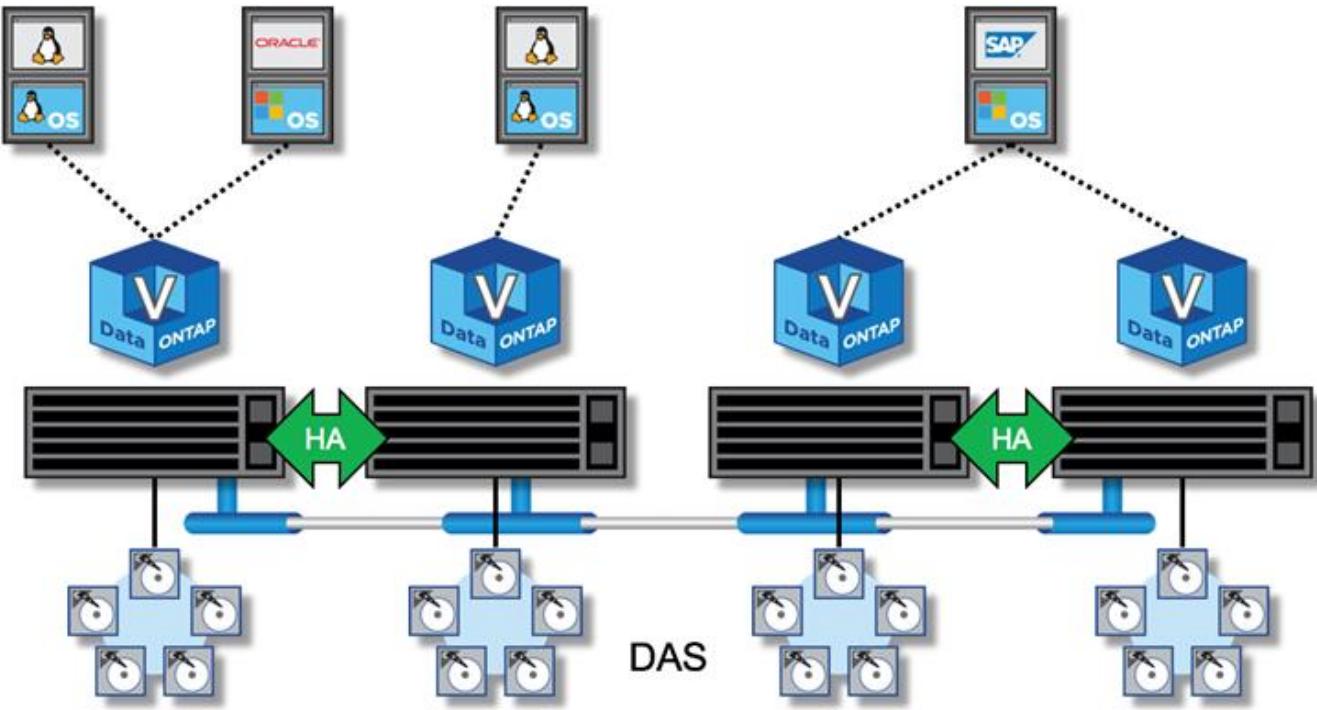
下圖表示了這兩種架構。

帶有遠端調解器並使用本地連接存儲的雙節點ONTAP Select集群



雙節點ONTAP Select叢集由一個 HA 對和一個調解器組成。在 HA 對中，每個叢集節點上的資料聚合都會同步鏡像，因此即使發生故障轉移，也不會遺失資料。

使用本地連接儲存的四節點ONTAP Select叢集



- 四節點ONTAP Select集群由兩個 HA 對組成。六節點和八節點群集分別由三個和四個 HA 對組成。在每個 HA 對中，每個叢集節點上的資料聚合都會同步鏡像，因此即使發生故障轉移，也不會遺失資料。
- 使用 DAS 儲存時，實體伺服器上只能存在一個ONTAP Select實例。ONTAPONTAP Select需要對系統的本機 RAID 控制器進行非共用訪問，並且旨在管理本機連接的磁碟，而如果沒有與儲存裝置的實體連接，則無法實現這一點。

## 雙節點 HA 與多節點 HA

與FAS陣列不同，HA 對中的ONTAP Select節點僅透過 IP 網路進行通訊。這意味著 IP 網路存在單點故障 (SPOF)，因此，防止網路分區和腦裂情況成為設計的重要方面。多節點叢集可以承受單節點故障，因為叢集仲裁可以由三個或更多倅存節點建立。雙節點叢集依賴ONTAP Deploy 虛擬機器所託管的調解器服務來實現相同的結果。

ONTAP Select節點和ONTAP Deploy 中介服務之間的心跳網路流量極小且具有彈性，因此ONTAP Deploy VM 可以託管在與ONTAP Select雙節點叢集不同的資料中心。

i ONTAP Deploy 虛擬機器在充當雙節點叢集的調解器時，將成為該叢集不可或缺的一部分。如果調解器服務不可用，雙節點叢集將繼續提供數據，但ONTAP Select叢集的儲存故障轉移功能將被停用。因此，ONTAP Deploy 調解器服務必須與 HA 對中的每個ONTAP Select節點保持持續通訊。為了確保叢集仲裁正常運行，最低頻寬要求為 5Mbps，最大往返時間 (RTT) 延遲要求為 125 毫秒。

如果充當調解器的ONTAP Deploy 虛擬機器暫時或可能永久無法使用，則可以使用輔助ONTAP Deploy 虛擬機器來還原雙節點叢集仲裁。這會導致新的ONTAP Deploy 虛擬機器無法管理ONTAP Select節點，但可以成功參與叢集仲裁演算法。ONTAPONTAP Select節點與ONTAP Deploy 虛擬機器之間的通訊是透過 IPv4 上的 iSCSI 協定完成的。ONTAPONTAP Select節點管理 IP 位址是啟動器，ONTAP Deploy 虛擬機器 IP 位址是目標。因此，在建立雙節點叢集時，節點管理 IP 位址無法支援 IPv6 位址。在建立雙節點叢集時，系統會自動建立ONTAP Deploy 託管的郵件磁碟，並將其封鎖為正確的ONTAP Select節點管理 IP 位址。整個配置在設定過程中自動執行，無需進一步的管理操作。建立叢集的ONTAP Deploy 實例是該叢集的預設調解器。

如果必須變更原始調解器位置，則需要執行管理操作。即使原始ONTAP Deploy 虛擬機器遺失，也可以復原叢集仲裁。但是，NetApp建議您在每個雙節點叢集實例化後備份ONTAP Deploy 資料庫。

## 雙節點 HA 與雙節點延伸 HA (MetroCluster SDS)

可以將雙節點主動/主動 HA 叢集擴展到更遠的距離，並可能將每個節點放置在不同的資料中心。雙節點群集和雙節點延伸群集（也稱為MetroCluster SDS）之間的唯一區別在於節點之間的網路連接距離。

雙節點叢集定義為兩個節點位於相同資料中心且相距 300 公尺以內的叢集。通常，兩個節點都具有到同一網路交換器或一組交換器間連結 (ISL) 網路交換器的上行鏈路。

雙節點MetroCluster SDS 是指節點物理上相距超過 300 公尺（不同房間、不同建築物和不同資料中心）的群集。此外，每個節點的上行鏈路連接分別連接到單獨的網路交換器。MetroCluster MetroCluster 不需要專用硬體。但是，該環境應滿足延遲（RTT 最大 5 毫秒，抖動最大 5 毫秒，總計 10 毫秒）和物理距離（最大 10 公里）的要求。

MetroCluster SDS 是一項進階功能，需要 Premium 授權或 Premium XL 授權。Premium 授權支援建立中小型虛擬機器以及 HDD 和 SSD 媒體。PremiumXL 許可證還支援建立 NVMe 驅動器。



MetroCluster SDS 支援本地連接儲存 (DAS) 和共用儲存 (vNAS)。請注意，由於ONTAP Select虛擬機器與共用儲存之間的網路，vNAS 配置通常具有較高的固有延遲。MetroClusterSDS 配置必須在節點之間提供最多 10 毫秒的延遲，其中包括共用儲存延遲。換句話說，僅測量 Select 虛擬機器之間的延遲是不夠的，因為對於這些配置來說，共用儲存延遲不可忽略。

## ONTAP Select HA RSM 和鏡像聚合

使用 RAID SyncMirror (RSM)、鏡像聚合和寫入路徑防止資料遺失。

### 同步複製

ONTAP高可用性 (HA) 模型建立在高可用性合作夥伴的概念之上。ONTAPONTAP Select將此架構擴展到非共用商用伺服器領域，利用ONTAP中提供的 RAID SyncMirror (RSM) 功能在叢集節點之間複製資料區塊，從而在高可用性 (HA) 對中提供兩個使用者資料副本。

具有調解器的雙節點叢集可以跨越兩個資料中心。有關更多信息，請參閱["雙節點擴展 HA \(MetroCluster SDS\) 最佳實踐"](#)。

### 鏡像聚合

ONTAP Select叢集由 2 到 8 個節點組成。每個 HA 對包含兩個使用者資料副本，透過 IP 網路跨節點同步鏡像。此鏡像對使用者透明，並且是資料聚合的屬性，會在資料聚合建立過程中自動配置。

ONTAP Select叢集中的所有聚合都必須進行鏡像，以便在發生節點故障轉移時確保資料可用性，並在發生硬體故障時避免出現單點故障 (SPOF)。ONTAPONTAP Select叢集中的聚合基於 HA 對中每個節點提供的虛擬磁碟構建，並使用以下磁碟：

- 一組本機磁碟（由目前ONTAP Select節點提供）
- 一組鏡像磁碟（由目前節點的 HA 夥伴提供）



用於建構鏡像聚合的本機磁碟和鏡像磁碟的大小必須相同。這些聚合稱為叢 0 和叢 1（分別表示本地鏡像對和遠端鏡像對）。實際的叢編號在您的安裝中可能有所不同。

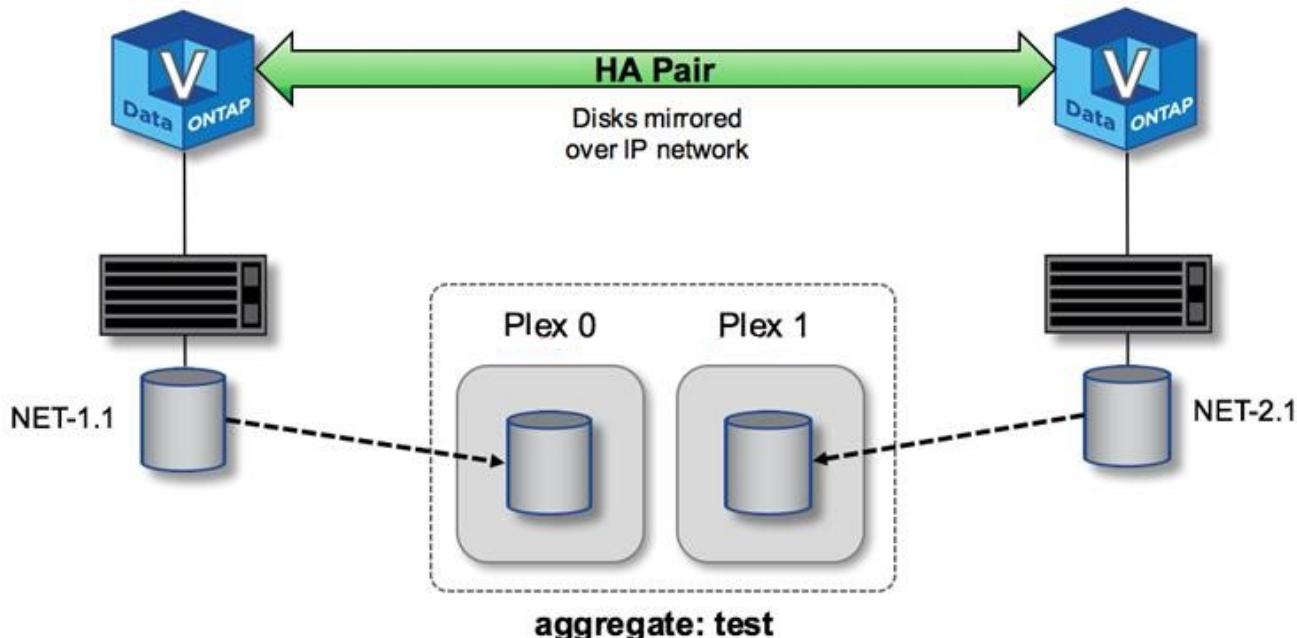
這種方法與標準ONTAP集群的工作方式有著根本的不同。這適用於ONTAP Select叢集中的所有根磁碟和資料磁碟。聚合包含資料的本機副本和鏡像副本。因此，包含 N 個虛擬磁碟的聚合可提供相當於 N/2 個磁碟的唯一存儲，因為第二個資料副本位於其自己的唯一磁碟上。

下圖顯示了四節點ONTAP Select叢集中的 HA 對。此叢集中有一個聚合（測試），它使用來自兩個 HA 配對節點的儲存。此資料聚合由兩組虛擬磁碟組成：一組本機磁碟，由ONTAP Select所屬叢集節點 (Plex 0) 提供；一組遠端磁碟，由故障轉移配對節點 (Plex 1) 提供。

Plex 0 是用於存放所有本機磁碟的儲存桶。Plex1 是用於存放鏡像磁碟（即負責儲存使用者資料第二個複製副本的磁碟）的儲存桶。擁有聚合的節點會將磁碟提供給 Plex 0，而該節點的 HA 配對節點會將磁碟提供給 Plex 1。

下圖中有一個包含兩個磁碟的鏡像聚合。此聚合的內容在兩個叢集節點之間進行鏡像，本地磁碟 NET-1.1 放置在 Plex 0 儲存桶中，遠端磁碟 NET-2.1 放置在 Plex 1 儲存桶中。在此範例中，聚合測試歸左側的叢集節點所有，並使用本機磁碟 NET-1.1 和 HA 夥伴鏡像磁碟 NET-2.1。

- ONTAP Select鏡像聚合\*



部署ONTAP Select叢集時，系統上的所有虛擬磁碟都會自動指派給正確的 Plex，而無需使用者執行任何與磁碟分配相關的額外步驟。這可以防止磁碟意外分配給錯誤的 Plex，並提供最佳的鏡像磁碟配置。

## 寫入路徑

叢集節點之間的資料區塊同步鏡像以及系統故障時不遺失資料的要求對傳入寫入作業在ONTAP Select叢集中傳播時所採用的路徑有重大影響。此過程包含兩個階段：

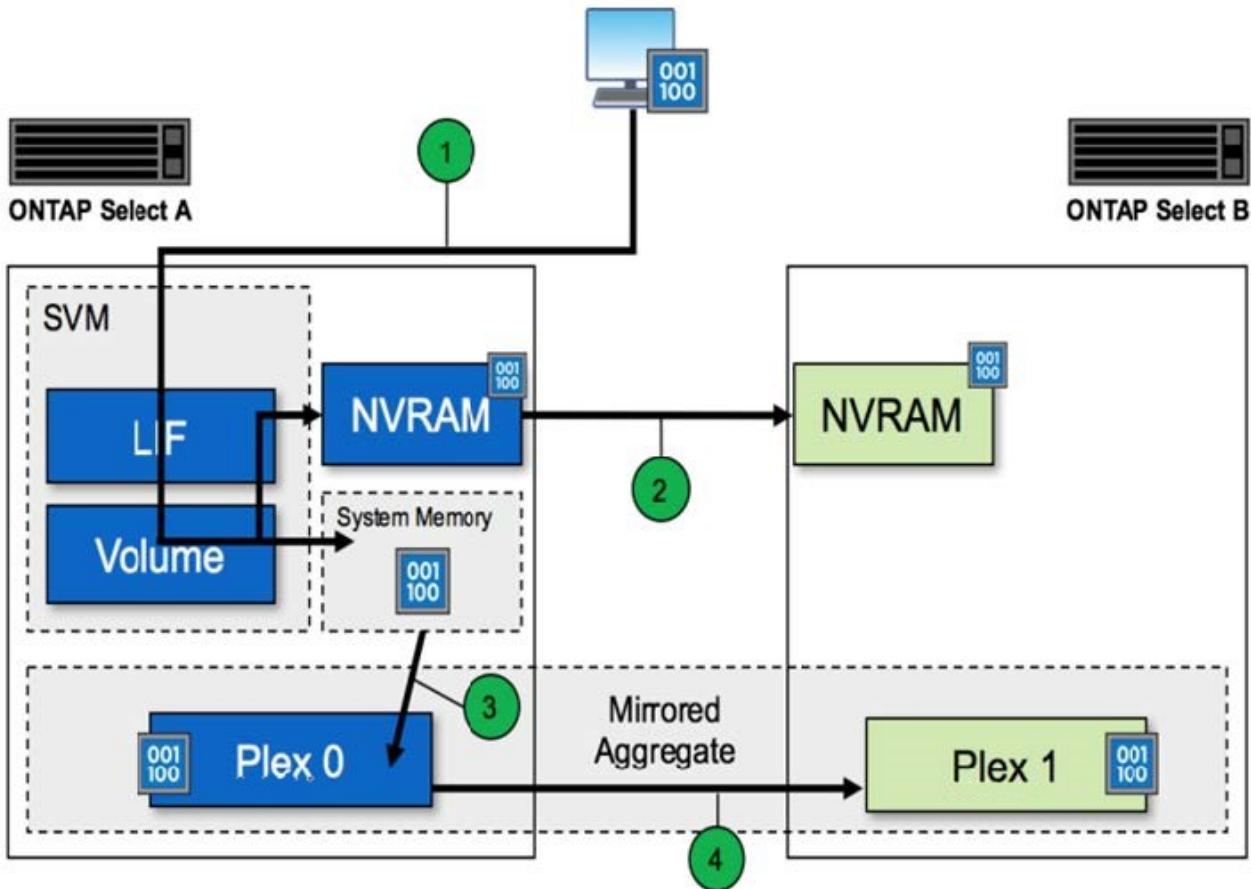
- 致謝
- 降級

對目標磁碟區的寫入作業透過資料 LIF 進行，並提交到ONTAP Select節點系統磁碟上的虛擬化NVRAM分割區，然後再向客戶端確認。在 HA 設定中，也會執行一個額外的步驟，因為這些NVRAM寫入作業會在確認之前立即鏡像到目標磁碟區擁有者的 HA 配對節點。此流程可確保在原始節點發生硬體故障時，HA 配對節點上的檔案系統保持一致。

將寫入內容提交至NVRAM後，ONTAP會定期將此分割區的內容移至對應的虛擬磁碟，此程序稱為降級轉儲。此過程僅在擁有目標磁碟區的叢集節點上發生一次，不會在 HA 配對節點上發生。

下圖顯示了傳入ONTAP Select節點的寫入請求的寫入路徑。

- ONTAP Select寫入路徑工作流程\*



傳入寫入確認包括以下步驟：

- 寫入透過ONTAP Select節點 A 擁有的邏輯介面進入系統。
- 寫入作業提交到節點 A 的NVRAM並鏡像到 HA 夥伴節點 B。
- 當兩個 HA 節點都出現 I/O 請求後，該請求就會被確認回客戶端。

ONTAP Select從NVRAM降級到資料聚合 (ONTAP CP) 包括以下步驟：

- 寫入操作從虛擬NVRAM轉入虛擬資料聚合。
- 鏡像引擎同步將區塊複製到兩個 plex。

# ONTAP Select HA 增強資料保護

高可用性 (HA) 磁碟心跳、HA 信箱、HA 心跳、HA 故障轉移和回饋功能可增強資料保護。

## 磁碟心跳

儘管ONTAP Select HA 架構充分利用了傳統FAS陣列的許多程式碼路徑，但也存在一些例外。其中一個例外是基於磁碟的心跳機制的實現，這是一種非基於網路的通訊方法，叢集節點使用這種方法來防止網路隔離導致裂腦行為。裂腦場景是群集分區的結果，通常由網路故障引起，導致群集兩端都認為對方已宕機並試圖接管群集資源。

企業級 HA 實施必須妥善處理此類情況。ONTAP透過一種基於磁碟的客製化心跳機制來實現這一點。ONTAP郵箱負責處理此任務，它是叢集節點用於傳遞心跳訊息的實體儲存位置。這有助於集群確定連接性，從而在發生故障轉移時確定仲裁。

在使用共享儲存 HA 架構的FAS陣列上，ONTAP透過以下方式解決裂腦問題：

- SCSI 持久預留
- 持久 HA 元數據
- 透過 HA 互連發送 HA 狀態

然而，在ONTAP Select叢集的無共用架構中，節點只能看到自己的本機存儲，而看不到 HA 配對節點的本機儲存。因此，當網路分區隔離 HA 對的每一側時，上述確定集群仲裁和故障轉移行為的方法將無法使用。

雖然現有的裂腦檢測和避免方法無法使用，但仍需要一種能夠適應無共享環境約束的調解方法。ONTAPONTAP Select進一步擴展了現有的郵箱基礎架構，使其能夠在網路分區時充當調解方法。由於共用儲存不可用，因此調解是透過 NAS 存取郵箱磁碟來完成的。這些磁碟分佈在整個叢集中，包括雙節點叢集中的調解器，並使用 iSCSI 協定。因此，叢集節點可以根據對這些磁碟的存取做出智慧故障轉移決策。如果一個節點可以存取其 HA 夥伴節點之外的其他節點的郵箱磁碟，則該節點很可能已啟動且運作正常。



解決叢集仲裁和裂腦問題的郵箱架構和基於磁碟的心跳方法是ONTAP Select多節點變體需要四個獨立節點或雙節點叢集的調解器的原因。

## HA郵箱發帖

HA 郵箱架構採用訊息發布模型。叢集節點會定期向叢集中所有其他郵件磁碟（包括中介節點）發布訊息，表示該節點已啟動並正在執行。在正常運作的叢集中，任何時間點，叢集節點上的單一郵箱磁碟都會收到來自所有其他叢集節點的訊息。

每個 Select 叢集節點都附加一個虛擬磁碟，專門用於共用郵箱存取。此磁碟被稱為中介郵箱磁碟，因為其主要功能是在發生節點故障或網路分割時充當叢集中介。此郵箱磁碟包含每個叢集節點的分割區，並由其他 Select 叢集節點透過 iSCSI 網路掛載。這些節點會定期將運作狀況發佈到郵件匣磁碟的對應分割區。使用分佈在整個叢集中的網路可存取郵箱磁碟，您可以透過可達性矩陣推斷節點的運作狀況。例如，叢集節點 A 和 B 可以向叢集節點 D 的郵箱傳送郵件，但不能傳送郵件給節點 C 的郵件。此外，叢集節點 D 無法向節點 C 的郵箱發送郵件，因此節點 C 很可能已關閉或網路隔離，應該被接管。

## HA 心跳

與NetApp FAS平台一樣，ONTAP Select會定期透過 HA 互連發送 HA 心跳訊息。在ONTAP Select叢集中，此操作透過 HA 夥伴節點之間的 TCP/IP 網路連線執行。此外，基於磁碟的心跳訊息會傳遞到所有 HA 郵箱磁碟，

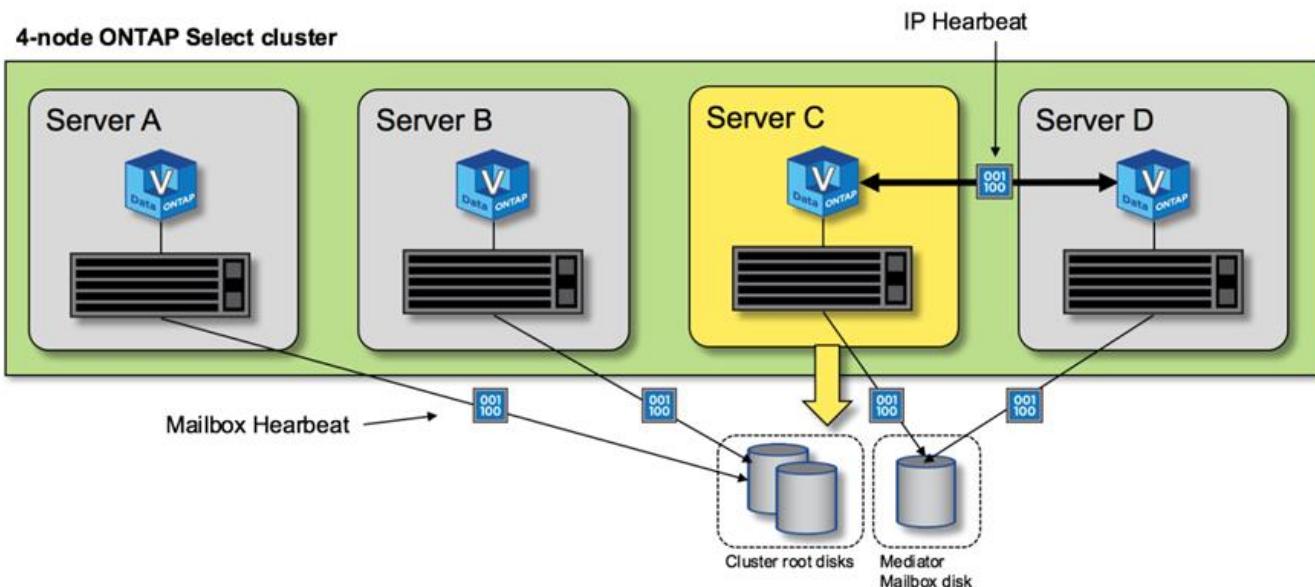
包括中介郵箱磁碟。這些訊息每隔幾秒鐘傳遞一次，並定期讀取。如此高的發送和接收頻率使ONTAP Select叢集能夠在大約 15 秒內檢測到 HA 故障事件，這與FAS平台上的可用時間視窗相同。當不再讀取心跳訊息時，將觸發故障轉移事件。

下圖從單一ONTAP Select叢集節點（節點 C）的角度顯示了透過 HA 互連和調解磁碟發送和接收心跳訊息的過程。



網路心跳透過 HA 互連傳送到 HA 夥伴節點 D，而磁碟心跳使用跨所有叢集節點 A、B、C 和 D 的郵件磁碟。

#### 四節點叢集中的 HA 心跳：穩定狀態



#### HA 故障轉移與恢復

在故障轉移作業期間，倖存節點將使用其 HA 夥伴節點資料的本機副本承擔其對等節點的資料服務責任。客戶端 I/O 可以繼續不間斷地運行，但必須先複製對此資料的更改，然後才能進行交還。請注意，ONTAP Select 不支援強制交還，因為這會導致儲存在倖存節點上的變更遺失。

重啟的節點重新加入叢集時，會自動觸發同步復原操作。同步恢復所需的時間取決於多種因素。這些因素包括必須複製的更改數量、節點之間的網路延遲以及每個節點上磁碟子系統的速度。同步恢復所需的時間可能會超過 10 分鐘的自動交還視窗。在這種情況下，需要在同步恢復後進行手動交還。您可以使用以下命令監控同步復原的進度：

```
storage aggregate status -r -aggregate <aggregate name>
```

## 版權資訊

Copyright © 2026 NetApp, Inc. 版權所有。台灣印製。非經版權所有人事先書面同意，不得將本受版權保護文件的任何部分以任何形式或任何方法（圖形、電子或機械）重製，包括影印、錄影、錄音或儲存至電子檢索系統中。

由 NetApp 版權資料衍伸之軟體必須遵守下列授權和免責聲明：

此軟體以 NETAPP 「原樣」提供，不含任何明示或暗示的擔保，包括但不限於有關適售性或特定目的適用性之擔保，特此聲明。於任何情況下，就任何已造成或基於任何理論上責任之直接性、間接性、附隨性、特殊性、懲罰性或衍生性損害（包括但不限於替代商品或服務之採購；使用、資料或利潤上的損失；或企業營運中斷），無論是在使用此軟體時以任何方式所產生的契約、嚴格責任或侵權行為（包括疏忽或其他）等方面，NetApp 概不負責，即使已被告知有前述損害存在之可能性亦然。

NetApp 保留隨時變更本文所述之任何產品的權利，恕不另行通知。NetApp 不承擔因使用本文所述之產品而產生的責任或義務，除非明確經過 NetApp 書面同意。使用或購買此產品並不會在依據任何專利權、商標權或任何其他 NetApp 智慧財產權的情況下轉讓授權。

本手冊所述之產品受到一項（含）以上的美國專利、國外專利或申請中專利所保障。

有限權利說明：政府機關的使用、複製或公開揭露須受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中的「技術資料權利 - 非商業項目」條款 (b)(3) 小段所述之限制。

此處所含屬於商業產品和 / 或商業服務（如 FAR 2.101 所定義）的資料均為 NetApp, Inc. 所有。根據本協議提供的所有 NetApp 技術資料和電腦軟體皆屬於商業性質，並且完全由私人出資開發。美國政府對於該資料具有非專屬、非轉讓、非轉授權、全球性、有限且不可撤銷的使用權限，僅限於美國政府為傳輸此資料所訂合約所允許之範圍，並基於履行該合約之目的方可使用。除非本文另有規定，否則未經 NetApp Inc. 事前書面許可，不得逕行使用、揭露、重製、修改、履行或展示該資料。美國政府授予國防部之許可權利，僅適用於 DFARS 條款 252.227-7015(b)（2014 年 2 月）所述權利。

## 商標資訊

NETAPP、NETAPP 標誌及 <http://www.netapp.com/TM> 所列之標章均為 NetApp, Inc. 的商標。文中所涉及的所有其他公司或產品名稱，均為其各自所有者的商標，不得侵犯。