高可用度架構
ONTAP Select

NetApp
May 21, 2024

目錄

高可用度架構

高可用度組態

探索高可用度選項、為您的環境選擇最佳HA組態。

雖然客戶開始將應用程式工作負載從企業級儲存設備移轉至以軟體為基礎的解決方案、並在市售硬體上執行、但對於恢復能力和容錯能力的期望和需求並未改變。HA解決方案提供零恢復點目標（RPO）、可保護客戶免受基礎架構堆疊中任何元件故障所造成的資料遺失。

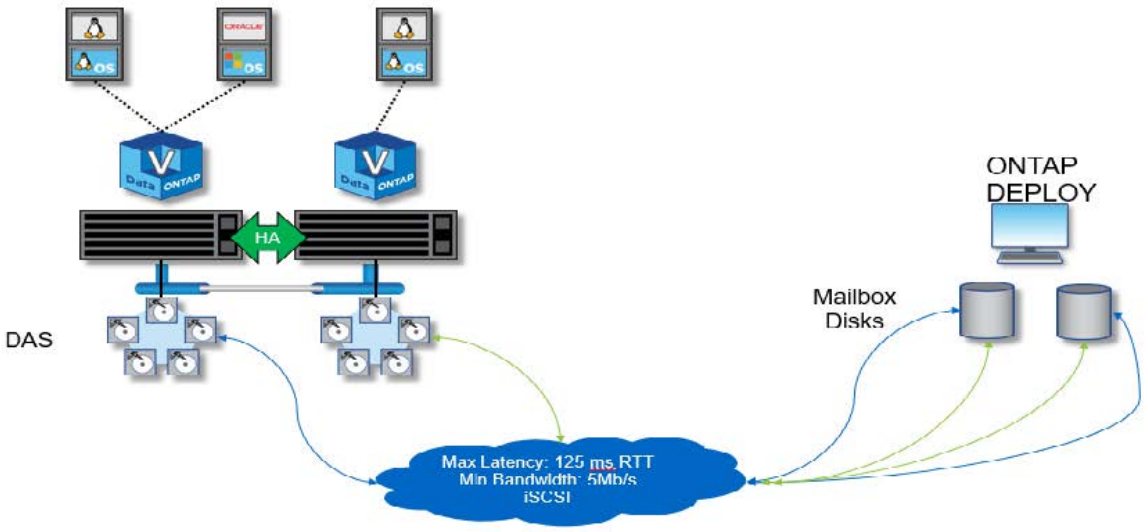
大部分的SDS市場都是以共享無儲存的概念為基礎、而軟體複寫則可在不同的儲存封閉環境中儲存多個使用者資料複本、提供資料恢復能力。利用由支援的同步複寫功能（RAID還原）、在這個前提下建置、以在叢集中儲存額外的使用者資料複本。ONTAP Select SyncMirror ONTAP這是發生在HA配對的內容中。每個HA配對都會儲存兩份使用者資料複本：一份儲存在本機節點所提供的儲存設備上、另一份儲存在HA合作夥伴所提供的儲存設備上。在一個叢集內ONTAP Select 、HA和同步複寫會相互連結、而且兩者的功能無法獨立分離或使用。因此、同步複寫功能僅適用於多節點產品。

在一個不同步叢集中、同步複寫功能是HA實作的功能、而非非非同步SnapMirror或還原複寫引擎的替代功能。ONTAP Select SnapVault無法從HA獨立使用同步複寫。

共有兩ONTAP Select 種功能：多節點叢集（四個、六個或八個節點）和雙節點叢集。雙節點ONTAP Select 的不二叢集的主要特色是使用外部中介服務來解決腦分割案例。此功能可作為預設的協調器、用於其所設定的所有雙節點HA配對。ONTAP

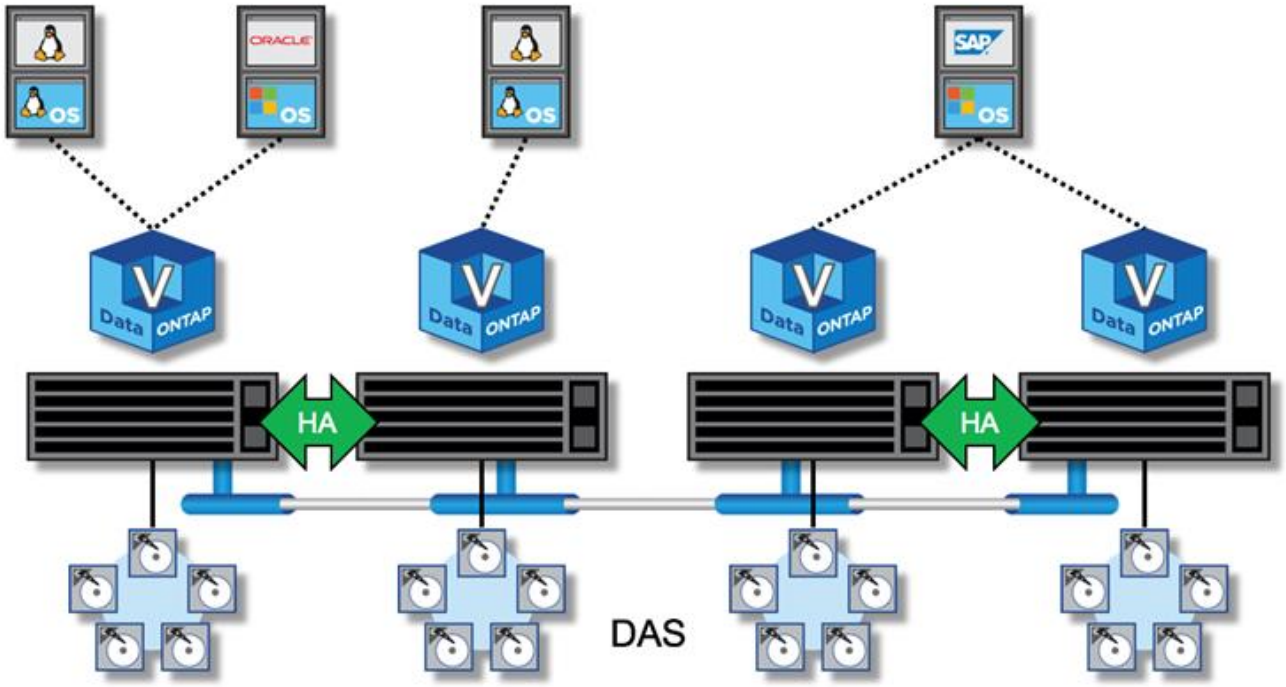
這兩種架構如下圖所示。

- ONTAP Select 使用遠端中介器和本機附加儲存設備*的雙節點版S叢 集



雙節點ONTAP Select 的叢集由一對HA配對和一位中介器組成。在HA配對中、每個叢集節點上的資料集合體都會同步鏡射、而且在發生容錯移轉時、資料不會遺失。

- ONTAP Select 使用本機附加儲存設備*的四節點效能叢集



- 四節點ONTAP Select 的叢集由兩個HA配對組成。六節點和八節點叢集分別由三對和四對HA組成。在每個HA配對中、每個叢集節點上的資料集合體都會同步鏡射、而且在發生容錯移轉時、資料不會遺失。
- 使用DAS儲存設備時、實體伺服器上只ONTAP Select 能有一個執行個體。不共享的系統本機RAID控制器存取權限、是專為管理本機附加磁碟而設計、如果沒有實體連線到儲存設備、就不可能做到這一點。ONTAP Select

雙節點**HA**與多節點**HA**的比較

與FAS 不一樣的是、ONTAP Select HA配對中的各個節點只能透過IP網路進行通訊。這表示IP網路是單點故障（SPOF）、防範網路分割區和腦分割案例、是設計的重要層面。多節點叢集可承受單一節點故障、因為叢集仲裁可由三個以上的正常節點建立。雙節點叢集仰賴ONTAP 由該解決方案代管的中介服務來達成相同的結果。

由於在各個節點與「支援不中斷」中介服務之間的活動訊號網路流量ONTAP Select 極少且彈性十分ONTAP 靈活、ONTAP 因此可將「部署虛擬機器」託管在ONTAP Select 不同於「支援不中斷」雙節點叢集的資料中心。

> 當用作該叢集的中介器時、支援功能可將虛擬機器整合為雙節點叢集的一部分。ONTAP如果無法使用中介服務、雙節點叢集會繼續提供資料、但ONTAP Select 會停用該叢集的儲存容錯移轉功能。因此ONTAP 、使用「零位元部署中介」服務時、必須持續與ONTAP Select HA配對中的每個節點通訊。為了讓叢集仲裁正常運作、至少需要5Mbps的頻寬和125ms的最大往返時間（RTT）延遲。

如果ONTAP 扮演中介角色的原地部署VM暫時或可能永久無法使用、可使用二線ONTAP 的原地部署VM來還原雙節點叢集仲裁。如此一來、新ONTAP 的S不到 位部署VM的組態就無法管理ONTAP Select 這些節點、但它成功參與了叢集仲裁演算法。透過透過IPV4使用iSCSI傳輸協定、即可在支援節點與支援資源中心之間進行通訊ONTAP Select 。ONTAP此為啟動器的靜態節點管理IP位址、而以此為目標的是部署VM IP位址。ONTAP Select ONTAP因此、建立雙節點叢集時、無法支援節點管理IP位址的IPv6位址。在建立雙節點叢集時、會自動建立及遮罩至適當的站台管理IP位址的子網路部署託管信箱磁碟。ONTAP ONTAP Select整個組態會在設定期間自動執行、不需要進一步的管理動作。建立叢集的程式集部署執行個體是該叢集的預設中介程式。ONTAP

如果必須變更原始的中介位置、則需要執行管理動作。即使原始ONTAP 的還原部署VM遺失、也能恢復叢集仲

裁。不過、NetApp建議您在ONTAP 每個雙節點叢集具現化之後、備份該還原部署資料庫。

雙節點HA與雙節點延伸HA（MetroCluster 不含雙節點）

您可以將雙節點主動/主動式HA叢集延伸至更遠的距離、並可能將每個節點放置在不同的資料中心。雙節點叢集與雙節點延伸叢集（也稱為MetroCluster 「效益SDS」）之間的唯一區別、是節點之間的網路連線距離。

雙節點叢集定義為叢集、兩個節點位於同一個資料中心、距離300公尺以內。一般而言、兩個節點都有上行鏈路可連至同一個網路交換器或一組交換器間連結（ISL）網路交換器。

雙節點MetroCluster 的不完整資料SDS定義為叢集、其節點實體分隔（不同的空間、不同的建築物和不同的資料中心）超過300公尺。此外、每個節點的上行鏈路連線都會連接至不同的網路交換器。不需要專用硬體即可實現此功能。MetroCluster不過、環境應符合延遲需求（RTT最長5毫秒、不穩定性最長5毫秒、總共10毫秒）和實體距離（最大10公里）。

MetroCluster SDS 是一項優質功能、需要 Premium 授權或 Premium XL 授權。Premium授權可支援建立中小型VM、以及HDD和SSD媒體。Premium XL 授權也支援 NVMe 磁碟機的建立。

> ⓘ 本地附加儲存設備（DAS）和共享儲存設備（vNAS）均支援使用支援的支援。MetroCluster請注意、VNAS組態通常會因為ONTAP Select 在整個過程中使用的是介於整個過程中的虛擬機器和共享儲存設備、而產生較高的內在延遲。包含共享儲存延遲在內的各個節點之間、均必須提供最多10毫秒的延遲。MetroCluster換句話說、只測量Select VM之間的延遲並不足夠、因為這些組態的共享儲存延遲並不微不足道。

HA RSM與鏡射Aggregate

使用RAID SyncMirror 功能（RSM2）、鏡射集合體和寫入路徑來防止資料遺失。

同步複寫

此功能是以HA合作夥伴的概念為基礎打造而成。ONTAP透過使用支援RAID功能（RSMs）、將此架構延伸至非共享的市售伺服器世界、此功能可在叢集節點之間複寫資料區塊、並在HA配對之間提供兩份使用者資料複本。ONTAP Select SyncMirror ONTAP

具有中介器的雙節點叢集可橫跨兩個資料中心。如需詳細資訊、請參閱一節 "雙節點延伸HA MetroCluster （簡稱「架構SDS」）最佳實務做法"。

鏡射Aggregate

一個由兩到八個節點組成的叢集。ONTAP Select每個HA配對都包含兩份使用者資料複本、透過IP網路在節點之間同步鏡射。此鏡射對使用者而言是透明的、而且是資料集合體的屬性、會在資料集合體建立程序期間自動設定。

在發生節點容錯移轉時、必須鏡射整個叢集中的所有集合體ONTAP Select 、以確保資料可用度、並避免發生硬體故障時出現SPOF。叢集中的Aggregate ONTAP Select 是從HA配對中的每個節點所提供的虛擬磁碟建置、並使用下列磁碟：

- 本機磁碟集（由目前ONTAP Select 的節點所提供）
- 鏡射磁碟集（由目前節點的HA合作夥伴提供）

用來建置鏡射Aggregate的本機磁碟和鏡射磁碟大小必須相同。這些集合體稱為plex 0和plex 1（分別表示本機和遠端鏡像配對）。實際的叢數在您的安裝中可能有所不同。

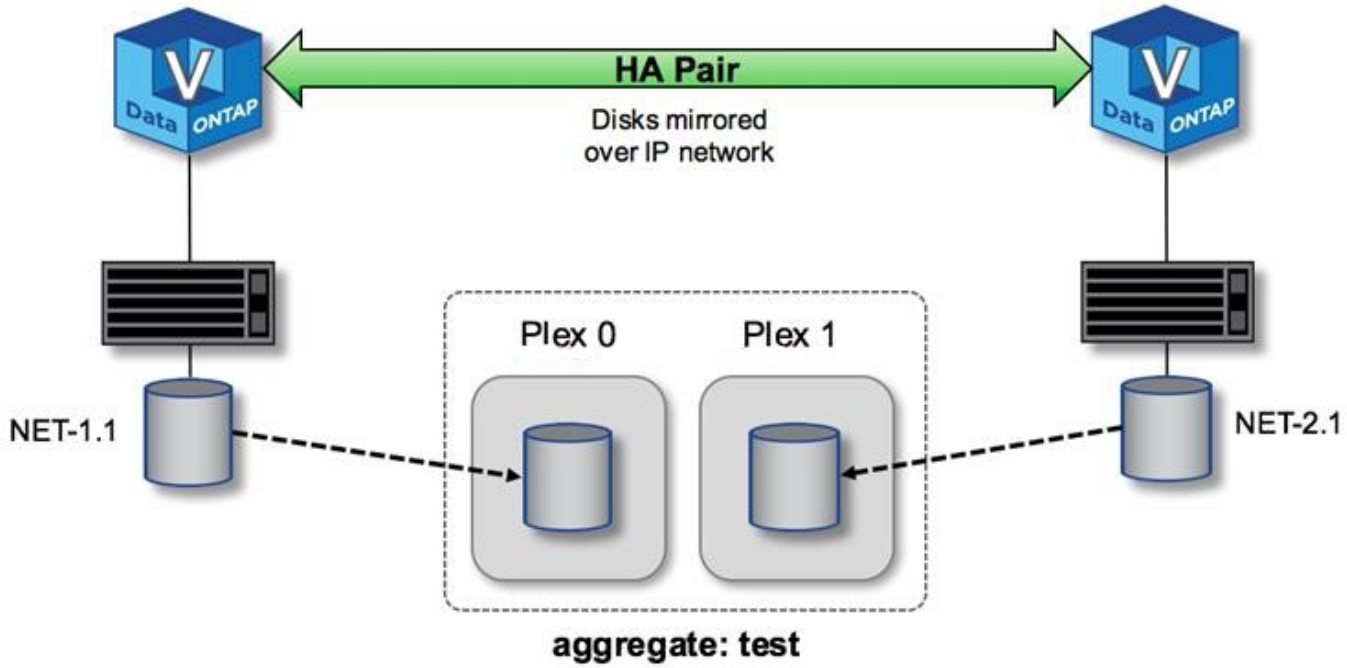
這種方法與標準ONTAP 的仰賴叢集運作方式有根本的不同。這適用於ONTAP Select 整個叢集內的所有根磁碟和資料磁碟。Aggregate同時包含本機和鏡射資料複本。因此，包含N個虛擬磁碟的集合體可提供不含2個磁碟的獨特儲存容量，因為第二個資料複本位於其專屬的磁碟上。

下圖顯示四節點ONTAP Select 的叢集內的HA配對。在此叢集中、是使用兩個HA合作夥伴儲存設備的單一集合體（測試）。此資料Aggregate由兩組虛擬磁碟組成：由ONTAP Select 故障轉移合作夥伴（Plex 1）所提供的實體叢集節點（Plex 0）所構成的本機磁碟集、以及遠端磁碟集。

Plex 0是存放所有本機磁碟的儲存區。Plex 1是存放鏡射磁碟的儲存區、或是負責儲存第二個使用者資料複本的磁碟。擁有Aggregate的節點會將磁碟貢獻給Plex 0、而該節點的HA合作夥伴則會將磁碟貢獻給Plex 1。

下圖中有一個鏡射Aggregate、其中包含兩個磁碟。此Aggregate的內容會鏡射至我們的兩個叢集節點、並將本機磁碟NET-1.1放入Plex 0儲存區、並將遠端磁碟NET-2.1放入Plex 1儲存區。在此範例中、Aggregate測試由左側的叢集節點擁有、並使用本機磁碟NET-1.1和HA合作夥伴鏡射磁碟NET-2.1。

*鏡射Aggregate * ONTAP Select



部署一個叢集時ONTAP Select 、系統上的所有虛擬磁碟都會自動指派給正確的叢集、不需要使用者就磁碟指派採取額外步驟。如此可防止意外將磁碟指派給不正確的叢、並提供最佳的鏡射磁碟組態。

寫入路徑

在叢集節點之間同步鏡射資料區塊、以及在系統故障時不需遺失資料、對於傳入寫入透過ONTAP Select 叢集傳播時所採用的路徑、會造成重大影響。此程序包含兩個階段：

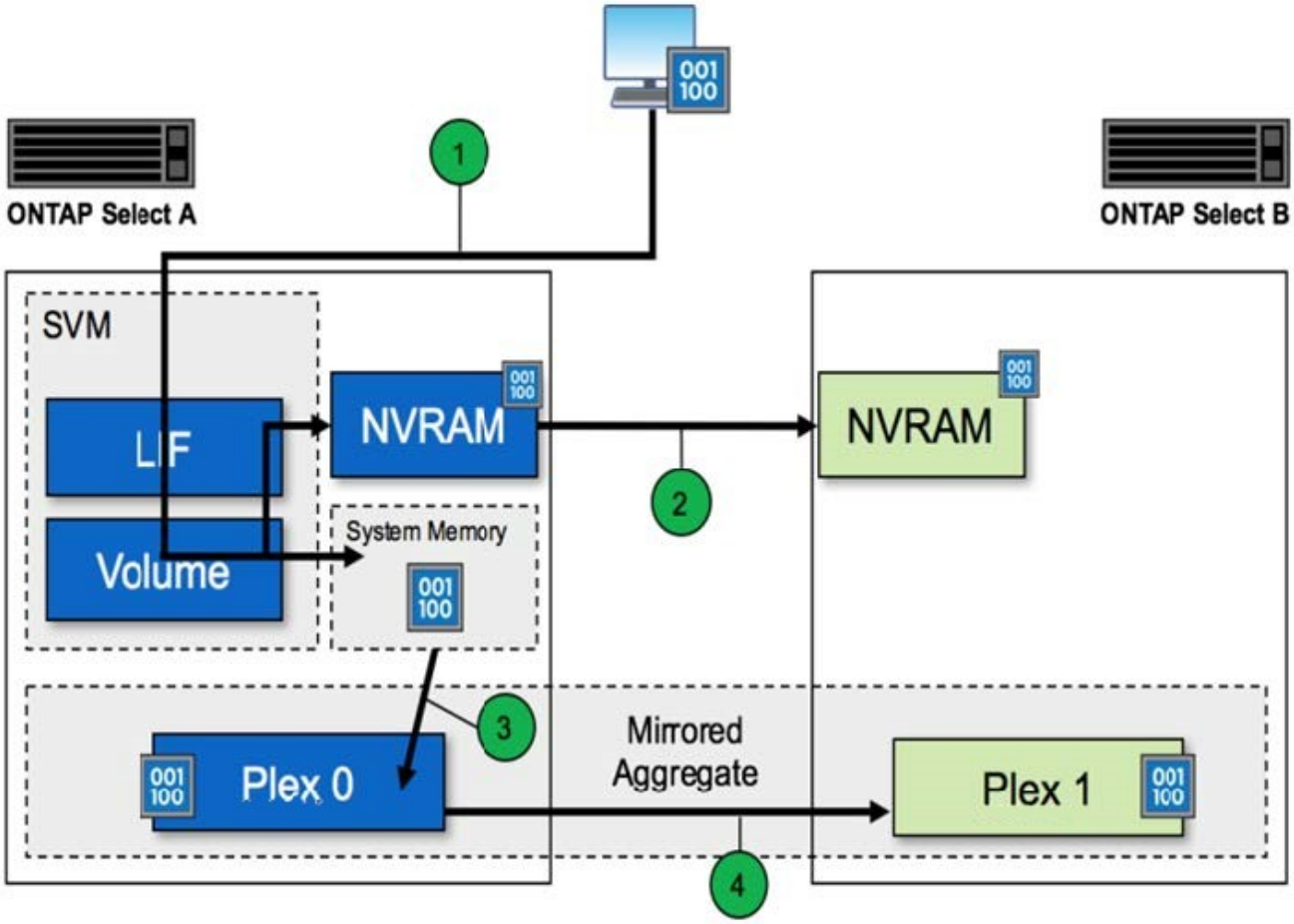
- 確認
- 減少需求

寫入目標磁碟區的作業會透過資料LIF進行、並將寫入到虛擬化的NVRAM分割區、此分割區會出現在ONTAP Select 節點的系統磁碟上、然後再確認回用戶端。在HA組態上、會執行額外的步驟、因為這些NVRAM寫入作業會在確認之前、立即鏡射到目標磁碟區擁有者的HA合作夥伴。如果原始節點發生硬體故障、此程序可確保HA合作夥伴節點上的檔案系統一致性。

將寫入作業提交至NVRAM後、ONTAP 將此分割區的內容定期移至適當的虛擬磁碟、這是稱為「減少磁碟空間」的程序。此程序只會在擁有目標Volume的叢集節點上執行一次、而不會發生在HA合作夥伴上。

下圖顯示傳入寫入要求至ONTAP Select 某個節點的寫入路徑。

不寫入路徑工作流程 ONTAP Select



傳入寫入認可包括下列步驟：

- 寫入會透過ONTAP Select 由節點A擁有的邏輯介面進入系統
- 寫入作業會提交至節點A的NVRAM、並鏡射至HA合作夥伴節點B
- 在兩個HA節點上都有I/O要求之後、該要求會被確認回用戶端。

從NVRAM降級至資料集合體（RACP）包括下列步驟：ONTAP Select ONTAP

- 寫入作業會從虛擬NVRAM移轉至虛擬資料Aggregate。
- 鏡射引擎會將區塊同步複寫到兩個叢集中。

HA其他詳細資料

HA磁碟心跳轉、HA信箱、HA心跳轉、HA容錯移轉及還原等功能、可強化資料保護。

磁碟心跳

雖然這個架構運用了許多傳統的指令碼路徑、但仍存在一些例外情況。ONTAP Select FAS其中一項例外是執行磁碟型心率、這是叢集節點用來防止網路隔離造成腦分割行為的非網路型通訊方法。分割腦部案例是叢集分割的結果、通常是網路故障所造成、因此每一方都認為另一方已停機、並嘗試接管叢集資源。

企業級HA實作必須妥善處理這類案例。透過客製化的磁碟型心跳法來達成此目標。ONTAP這是HA信箱的工作、這是叢集節點用來傳遞活動訊號訊息的實體儲存設備位置。這有助於叢集判斷連線能力、因此可在容錯移轉時定義仲裁。

在使用共享儲存HA架構的不同時陣列上、用下列方法解決大腦分裂問題：FAS ONTAP

- SCSI持續保留
- 持續HA中繼資料
- 透過HA互連傳送HA狀態

然而ONTAP Select 、在不共用的架構中、節點只能看到自己的本機儲存設備、而無法看到HA合作夥伴的儲存設備。因此、當網路分割區隔離HA配對的每一端時、就無法使用先前的判斷叢集仲裁和容錯移轉行為的方法。

雖然無法使用現有的分割腦偵測和避免方法、但仍需要一種協調方法、這種方法符合共享無事環境的限制。利用NetApp技術、可進一步擴充現有的信箱基礎架構、在網路分割的情況下、也能做為中繼方法。ONTAP Select由於共享儲存設備無法使用、因此透過NAS存取信箱磁碟來完成中繼作業。這些磁碟會使用iSCSI傳輸協定散佈到整個叢集、包括雙節點叢集中的中介器。因此、叢集節點可根據對這些磁碟的存取來做出智慧型容錯移轉決策。如果某個節點可以存取HA合作夥伴以外其他節點的信箱磁碟、則可能正常運作。

> 信箱架構和以磁碟為基礎的核心跳轉方法來解決叢集仲裁和腦分割問題、是ONTAP Select 導致多節點變體需要四個獨立節點、或是兩節點叢集的中介器的原因。

HA信箱張貼

HA信箱架構使用訊息POST模式。叢集節點會以重複的時間間隔、將訊息張貼到叢集內的所有其他信箱磁碟、包括中介器、說明節點已啟動且正在執行。在健全的叢集內、叢集節點上的單一信箱磁碟會在任何時間點張貼來自所有其他叢集節點的訊息。

附加至每個Select叢集節點的虛擬磁碟是專供共用信箱存取使用的磁碟。此磁碟稱為中介信箱磁碟、因為它的主要功能是在節點故障或網路分割時做為叢集中繼的方法。此信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。這些節點會定期將健全狀況狀態張貼到信箱磁碟的適當分割區。使用分散於整個叢集的網路存取信箱磁碟、可讓您透過可到達性對照表來推斷節點健全狀況。例如、叢集節點A和B可以POST至叢集節點D的信箱、但不能POST至節點C的信箱此外、叢集節點D無法POST至節點C的信箱、因此節點C可能已停機或是網路已隔離、因此應該接管。

HA令人心碎

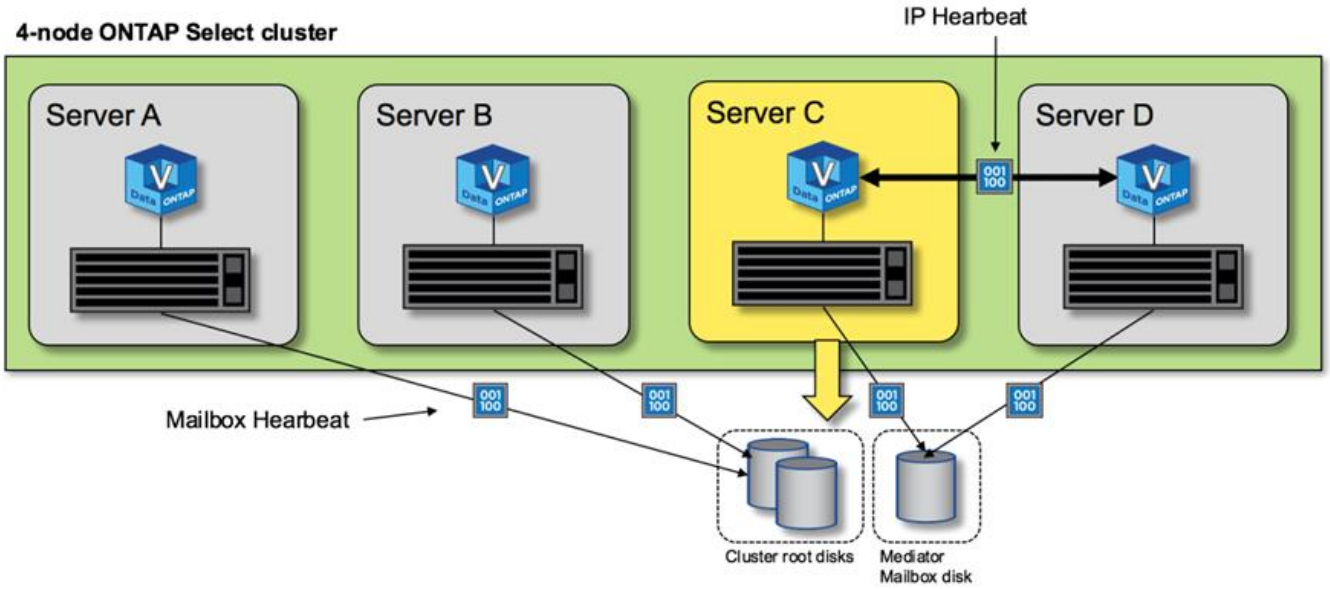
如同NetApp FAS 的支援平台、ONTAP Select VMware會定期透過HA互連傳送HA活動訊息。在這個叢集內ONTAP Select 、這是透過HA合作夥伴之間存在的TCP/IP網路連線來執行。此外、磁碟型的活動訊號訊息會傳遞至所有HA信箱磁碟、包括中介信箱磁碟。這些訊息會每隔幾秒傳遞一次、並定期讀取。發送和接收這些訊

號的頻率可讓ONTAP Select 故障叢集在約15秒內偵測HA故障事件、FAS 而這個視窗可在各個平台上使用。當不再讀取活動訊息時、會觸發容錯移轉事件。

下圖顯示從單ONTAP Select 一的故障叢集節點C的觀點、透過HA互連磁碟和中介磁碟傳送和接收訊號訊息的程序

網路訊號會透過HA互連傳送至HA合作夥伴節點D、而磁碟訊號則會在所有叢集節點、A、B、C和D之間使用信箱磁碟

在四節點叢集中執行**HA**資料中心：穩定狀態



HA容錯移轉與還原

在容錯移轉作業期間、正常運作的節點會使用HA合作夥伴資料的本機複本、承擔對等節點的資料服務責任。用戶端I/O可以不中斷地繼續運作、但必須先複寫此資料的變更、才能進行還原。請注意ONTAP Select 、由於這會導致儲存在存續節點上的變更遺失、所以不支援強制恢復。

重新開機的節點重新加入叢集時、會自動觸發同步處理回復作業。同步回溯所需的時間取決於多項因素。這些因素包括必須複寫的變更數目、節點之間的網路延遲、以及每個節點上磁碟子系統的速度。同步處理回饋所需的時間可能超過自動回饋10分鐘。在這種情況下、需要在同步回溯後手動恢復。您可以使用下列命令來監控同步處理回復的進度：

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