Clúster de almacenamiento vSphere Metro con ONTAP

Enterprise applications

NetApp
May 03, 2024

Tabla de contenidos

Clúster de almacenamiento vSphere Metro con ONTAP . . . 1
Clúster de almacenamiento vSphere Metro con ONTAP . . . 1
Descripción general de la solución de VMware vSphere . . . 4
Directrices de implementación y diseño de VMSC . . . 9
Resiliencia para eventos planificados y no planificados . . . 20
Escenarios de fallo para VMSC con MCC . . . 21

Clúster de almacenamiento vSphere Metro con ONTAP

Clúster de almacenamiento vSphere Metro con ONTAP

El hipervisor vSphere líder del sector de VMware se puede poner en marcha como un clúster ampliado conocido como vSphere Metro Storage Cluster (VMSC).

Las soluciones VMSC son compatibles con NetApp® MetroCluster™ y SnapMirror Active Sync (anteriormente conocido como continuidad empresarial de SnapMirror o SMBC) y proporcionan continuidad empresarial avanzada si uno o más dominios de fallo sufren una interrupción total. La resistencia a los diferentes modos de fallo depende de las opciones de configuración que elija.

Soluciones de disponibilidad continua para entornos vSphere

La arquitectura ONTAP es una plataforma de almacenamiento flexible y escalable que proporciona servicios SAN (FCP, iSCSI y NVMe-oF) y NAS (NFS v3 y v4,1) para almacenes de datos. Los sistemas de almacenamiento NetApp AFF, ASA y FAS utilizan el sistema operativo ONTAP para ofrecer protocolos adicionales para acceso al almacenamiento invitado, como S3 y SMB/CIFS.

NetApp MetroCluster utiliza la función de alta disponibilidad (conmutación por error de controladora o director financiero) de NetApp para proteger frente a fallos de controladora. También incluye tecnología SyncMirror local, recuperación tras fallos en clúster en caso de desastre (conmutación por error de controladora bajo demanda o CFOD), redundancia de hardware y separación geográfica para lograr altos niveles de disponibilidad. SyncMirror refleja de forma síncrona los datos en las dos mitades de la configuración de MetroCluster mediante la escritura de los datos en dos plexes: El plex local (en la bandeja local) que sirve los datos de forma activa y el plex remoto (en la bandeja remota) normalmente no ofrece datos. La redundancia de hardware se pone en marcha para todos los componentes de MetroCluster, como las controladoras, el almacenamiento, los cables, los switches (utilizados con Fabric MetroCluster) y los adaptadores.

La sincronización activa de SnapMirror de NetApp ofrece protección granular de almacenes de datos con protocolos SAN FCP e iSCSI, lo que permite proteger de forma selectiva solo las cargas de trabajo de alta prioridad. Ofrece acceso activo-activo tanto a sitios locales como remotos, a diferencia de NetApp MetroCluster, que es una solución activa-en espera. En la actualidad, la sincronización activa es una solución asimétrica en la que se prefiere un lado sobre el otro, lo que proporciona un mejor rendimiento. Esto se logra mediante la funcionalidad ALUA (acceso asimétrico de unidad lógica), que informa automáticamente al host ESXi qué prefieren las controladoras. Sin embargo, NetApp ha anunciado que la sincronización activa pronto permitirá un acceso totalmente simétrico.

Para crear un clúster HA/DRS de VMware en dos sitios, los hosts ESXi se usan y gestionan mediante una instancia de vCenter Server Appliance (VCSA). Las redes de gestión de vSphere, vMotion® y máquinas virtuales están conectadas a través de una red redundante entre los dos sitios. El servidor vCenter que gestiona el clúster HA/DRS puede conectarse a los hosts ESXi en ambos sitios y se debe configurar mediante vCenter HA.

Consulte "¿Cómo se crean y configuran clústeres en vSphere Client" Para configurar una alta disponibilidad de vCenter.

También debe consultar "Prácticas recomendadas para VMware vSphere Metro Storage Cluster".

¿Qué es vSphere Metro Storage Cluster?

VSphere Metro Storage Cluster (VMSC) es una configuración certificada que protege las máquinas virtuales (VM) y los contenedores frente a fallos. Esto se logra mediante el uso de conceptos de almacenamiento extendidos junto con clústeres de hosts ESXi, que se distribuyen en diferentes dominios de fallo, como bastidores, edificios, campus o incluso ciudades. Las tecnologías de almacenamiento de sincronización activa de NetApp MetroCluster y SnapMirror se usan para proporcionar una protección con un objetivo de punto de recuperación=0 o cerca del objetivo de punto de recuperación=0 respectivamente en los clústeres de hosts. La configuración de VMSC está diseñada para garantizar que los datos estén siempre disponibles incluso en caso de que falle un «sitio» completo, físico o lógico. Un dispositivo de almacenamiento que forme parte de la configuración de VMSC debe estar certificado tras someterse a un proceso de certificación VMSC exitoso. Todos los dispositivos de almacenamiento admitidos se pueden encontrar en la "Guía de compatibilidad de almacenamiento de VMware".

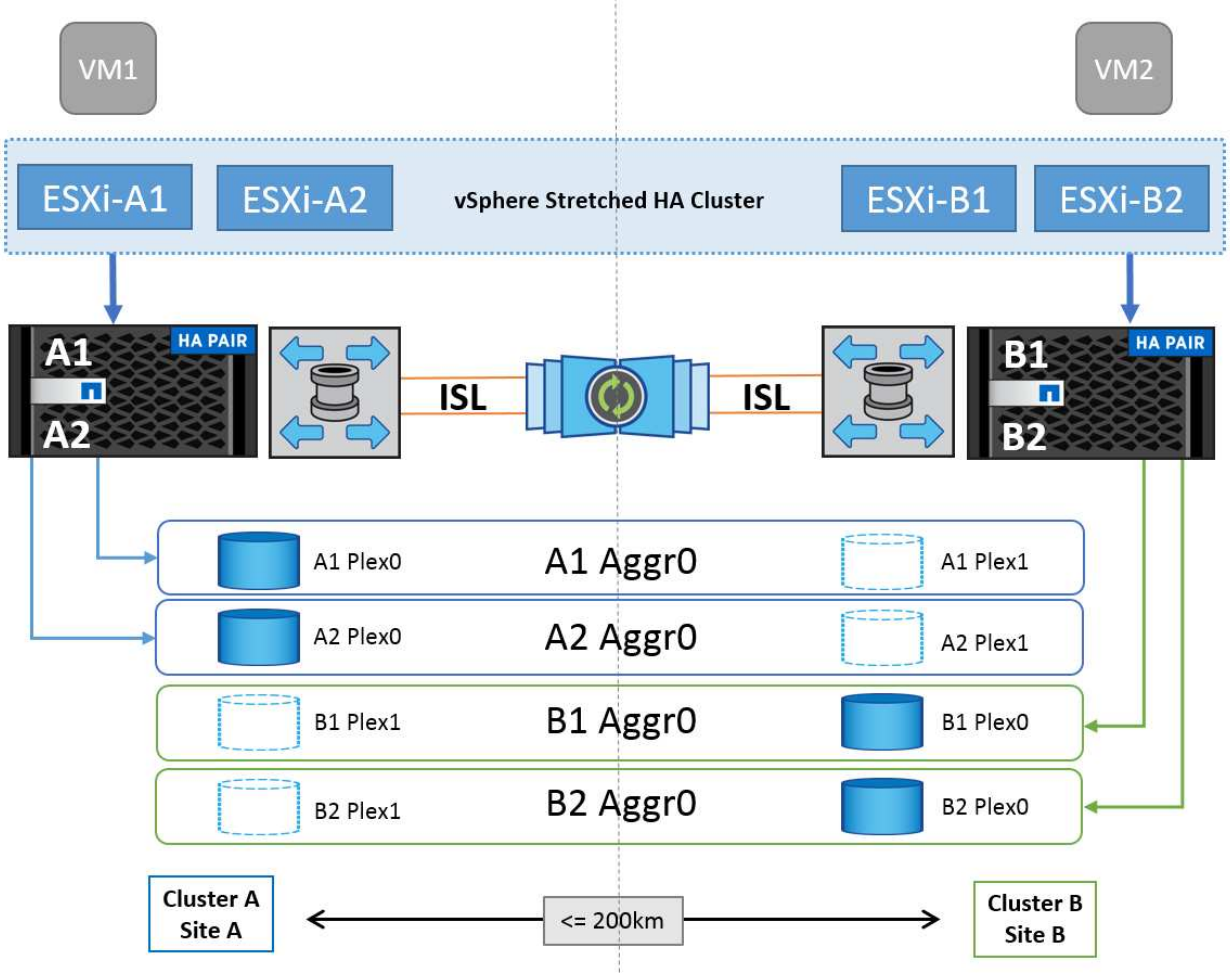
Si desea obtener más información sobre las directrices de diseño para vSphere Metro Storage Cluster, consulte la siguiente documentación:

- "Compatibilidad de VMware vSphere con NetApp MetroCluster"
- "Compatibilidad de VMware vSphere con Continuidad del negocio de SnapMirror de NetApp" (Ahora conocido como SnapMirror active sync)

Según las consideraciones de latencia, NetApp MetroCluster puede ponerse en marcha en dos configuraciones diferentes para utilizarlas con vSphere:

- Stretch MetroCluster
- Fabric MetroCluster

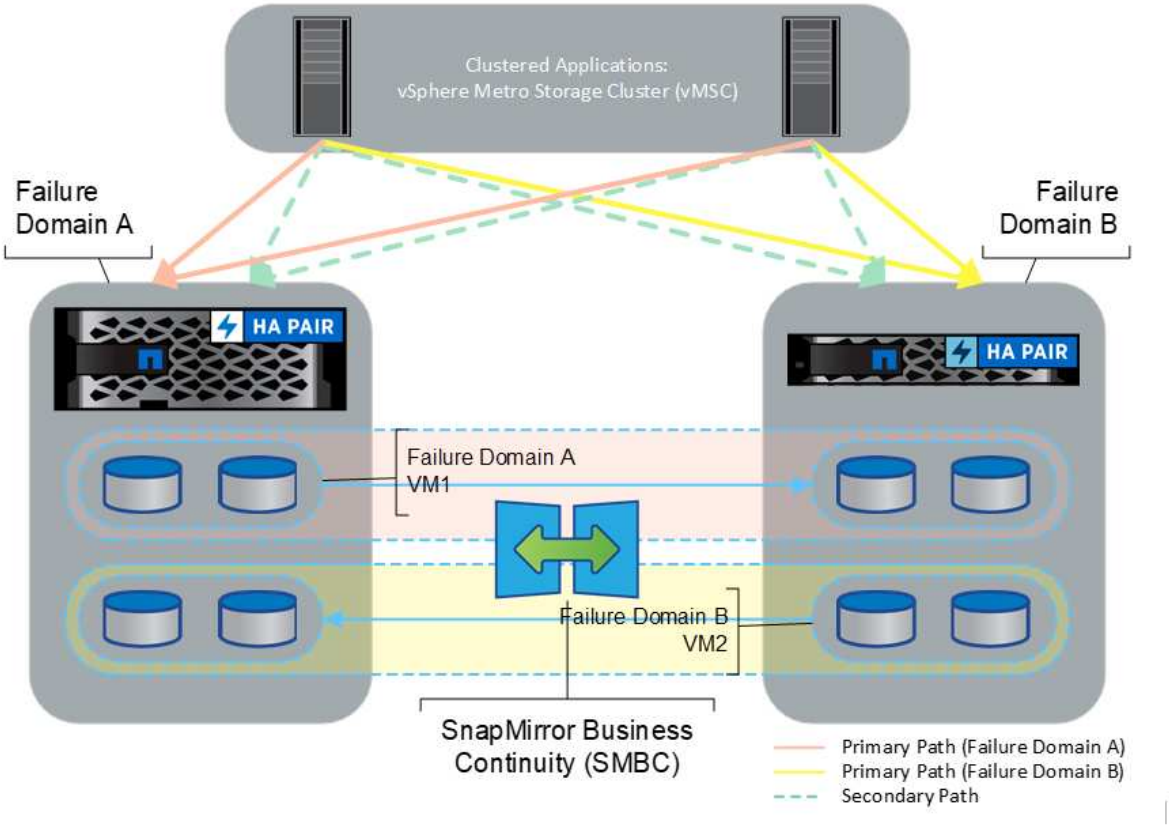
A continuación se muestra un diagrama topológico de alto nivel de MetroCluster de ampliación.



Consulte "Documentación de MetroCluster" Para obtener información específica sobre diseño e implementación para MetroCluster.

SnapMirror Active Sync también se puede poner en marcha de dos formas distintas.

- Asimétrico
- Simétrico (vista previa privada en ONTAP 9.14.1)



Consulte "Documentos de NetApp" Para obtener información específica de diseño e puesta en marcha para SnapMirror, sincronización activa.

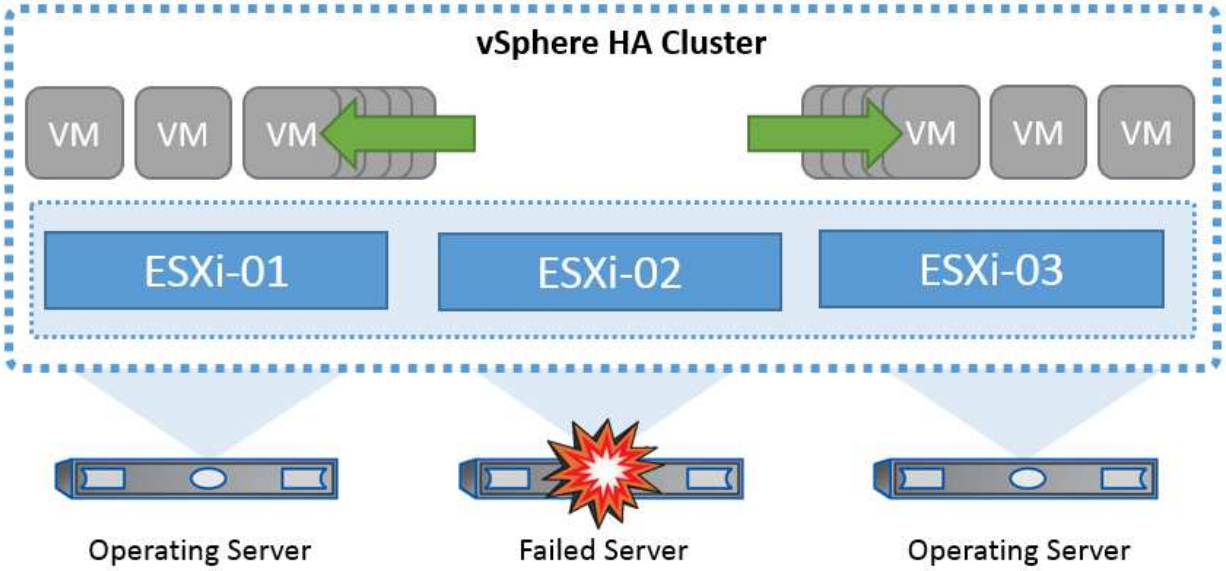
Descripción general de la solución de VMware vSphere

VCenter Server Appliance (VCSA) es el potente sistema de gestión centralizada y un panel único para vSphere que permiten a los administradores operar clústeres ESXi de forma eficiente. Facilita funciones clave como el aprovisionamiento de máquinas virtuales, la operación vMotion, alta disponibilidad, planificador de recursos distribuidos (DRS), Tanzu Kubernetes Grid, etc. Es un componente esencial en los entornos cloud de VMware y debe diseñarse teniendo en cuenta la disponibilidad del servicio.

Alta disponibilidad de vSphere

La tecnología de clúster de VMware agrupa servidores ESXi en pools de recursos compartidos para máquinas virtuales y proporciona vSphere High Availability (HA). VSphere HA proporciona una alta disponibilidad fácil de usar para aplicaciones que se ejecutan en máquinas virtuales. Cuando se habilita la función HA en el clúster, cada servidor ESXi mantiene la comunicación con otros hosts de modo que si algún host ESXi deja de responder o aísla, el clúster de alta disponibilidad puede negociar la recuperación de las máquinas virtuales que se estaban ejecutando en ese host ESXi entre los hosts supervivientes del clúster. Si se produce un fallo del sistema operativo invitado, vSphere HA reiniciará la máquina virtual afectada en el mismo servidor físico. La alta disponibilidad de vSphere permite reducir el tiempo de inactividad planificado, evitar tiempos de inactividad no planificados y recuperarse rápidamente de interrupciones.

Un clúster de alta disponibilidad de vSphere que recupera las máquinas virtuales del servidor con errores.



Es importante entender que VMware vSphere no tiene conocimientos de la sincronización activa de NetApp MetroCluster o SnapMirror y ve todos los hosts ESXi del clúster de vSphere como hosts elegibles para operaciones en clúster de alta disponibilidad en función de las configuraciones de afinidad de hosts y grupos de máquinas virtuales.

Detección de fallo de host

En cuanto se crea el clúster HA, todos los hosts del clúster participan en sus elecciones y uno de los hosts se convierte en un maestro. Cada esclavo realiza latidos de red al maestro y, a su vez, el maestro realiza latidos de red en todos los hosts esclavos. El host maestro de un clúster de alta disponibilidad de vSphere es responsable de detectar el fallo de hosts esclavos.

Según el tipo de error detectado, es posible que las máquinas virtuales que se ejecutan en los hosts deban conmutar al nodo de respaldo.

En un clúster de alta disponibilidad de vSphere se detectan tres tipos de fallos de host:

- Fallo: Un host deja de funcionar.
- Aislamiento: Un host se convierte en una red aislada.
- Partición: Un host pierde la conectividad de red con el host maestro.

El host maestro supervisa los hosts esclavos del cluster. Esta comunicación se realiza a través del intercambio de latidos de la red cada segundo. Cuando el host maestro deja de recibir estos latidos de un host esclavo, comprueba si hay vida activa del host antes de declarar que el host ha fallado. La comprobación de vida que realiza el host maestro es determinar si el host esclavo está intercambiando latidos con uno de los almacenes de datos. Además, el host maestro comprueba si el host responde a los ping ICMP enviados a sus direcciones IP de gestión para detectar si simplemente está aislado de su nodo maestro o completamente aislado de la red. Para ello, haga ping en la puerta de enlace predeterminada. Se pueden especificar manualmente una o varias direcciones de aislamiento para mejorar la fiabilidad de la validación de aislamiento.

Best Practice

NetApp recomienda especificar un mínimo de dos direcciones de aislamiento adicionales, y que cada una de estas direcciones sea local de sitio. Esto mejorará la fiabilidad de la validación del aislamiento.

Respuesta de aislamiento del host

Isolation Response es una configuración de vSphere HA que determina la acción que se activa en máquinas virtuales cuando un host de un clúster de vSphere HA pierde sus conexiones de red de gestión, pero continúa ejecutándose. Hay tres opciones para esta configuración, “Desactivado”, “Apagar y reiniciar VM” y “Apagar y reiniciar VM”.

“Apagar” es mejor que “Apagar”, que no vacía los cambios más recientes en el disco o las transacciones de confirmación. Si los equipos virtuales no se apagan en 300 segundos, se apagan. Para cambiar el tiempo de espera, utilice la opción avanzada das.isolationshutdowntimeout.

Antes de que HA inicie la respuesta de aislamiento, primero comprueba si el agente maestro HA de vSphere posee el almacén de datos que contiene los archivos de configuración de la máquina virtual. Si no es así, el host no activará la respuesta de aislamiento, porque no hay ningún maestro para reiniciar las máquinas virtuales. El host comprobará periódicamente el estado del almacén de datos para determinar si un agente de alta disponibilidad de vSphere que posee el rol maestro.

Best Practice

NetApp recomienda establecer la “Respuesta de aislamiento del host” en Desactivado.

Se puede producir una condición de cerebro dividido si un host se aísla o particiona desde el host maestro HA de vSphere y el maestro no puede comunicarse a través de los almacenes de datos de latido o mediante ping. El maestro declara que el host aislado está muerto y reinicia los equipos virtuales en otros hosts del cluster. Ahora existe una condición de cerebro dividido porque hay dos instancias de la máquina virtual en ejecución, solo una de las cuales puede leer o escribir los discos virtuales. Ahora se pueden evitar las condiciones del cerebro dividido configurando VM Component Protection (VMCP).

Protección de componentes de máquina virtual (VMCP)

Una de las mejoras de funciones de vSphere 6, relevante para la alta disponibilidad, es VMCP. VMCP proporciona protección mejorada contra todas las condiciones de pérdida permanente de dispositivos (APD) y de pérdida permanente de dispositivos (PDL) para bloques (FC, iSCSI, FCoE) y almacenamiento de archivos (NFS).

Pérdida permanente de dispositivo (PDL)

PDL es una condición que ocurre cuando un dispositivo de almacenamiento falla de forma permanente o se elimina de forma administrativa y no se espera que regrese. La cabina de almacenamiento NetApp emite un código de detección SCSI a ESXi que declara la pérdida permanente del dispositivo. En la sección Condiciones de fallo y Respuesta de VM de vSphere HA, puede configurar cuál debe ser la respuesta después de detectar una condición PDL.

Best Practice

NetApp recomienda configurar “Response for Datastore with PDL” en “**Apagar y reiniciar VMs**”. Cuando se detecta esta condición, una máquina virtual se reinicia instantáneamente en un host en buen estado dentro del clúster de alta disponibilidad de vSphere.

Todas las rutas hacia abajo (APD)

APD es una condición que se produce cuando el host vuelve inaccesible a un dispositivo de almacenamiento y no hay rutas disponibles a la cabina. ESXi considera que esto es un problema temporal con el dispositivo y espera que vuelva a estar disponible.

Cuando se detecta una condición de APD, se inicia un temporizador. Después de 140 segundos, la condición APD se declara oficialmente, y el dispositivo se marca como APD Time Out. Una vez transcurridos los 140 segundos, HA comenzará a contar el número de minutos especificado en el APD de retraso para failover de VM. Cuando transcurra el tiempo especificado, HA reiniciará los equipos virtuales afectados. Puede configurar VMCP para que responda de manera diferente si lo desea (Desactivado, Incidir eventos o Apagar y reiniciar VM).

Best Practice

NetApp recomienda configurar “Response for Datastore with APD” en “**Apagar y reiniciar VMs (conservative)**”.

Conservative hace referencia a la probabilidad de que la alta disponibilidad pueda reiniciar equipos virtuales. Cuando se establece en Conservador, HA solo reiniciará la VM que se ve afectada por el APD si sabe que otro host puede reiniciarla. En caso de agresividad, HA intentará reiniciar la máquina virtual incluso si no conoce el estado de los otros hosts. Esto puede provocar que las máquinas virtuales no se reinicien si no hay ningún host con acceso al almacén de datos en el que se encuentra.

Si el estado APD se resuelve y el acceso al almacenamiento se restaura antes de que se agote el tiempo de espera, HA no reiniciará innecesariamente la máquina virtual a menos que se configure explícitamente para ello. Si se desea una respuesta, incluso cuando el entorno se ha recuperado de la condición APD, la respuesta para la recuperación APD después del tiempo de espera APD debe configurarse para restablecer las máquinas virtuales.

Best Practice

NetApp recomienda configurar la respuesta para la recuperación de APD después del tiempo de espera de APD en Desactivado.

Implementación de VMware DRS para NetApp MetroCluster

VMware DRS es una función que agrega los recursos de host en un clúster y se usa principalmente para equilibrar cargas dentro de un clúster de una infraestructura virtual. VMware DRS calcula principalmente los recursos de la CPU y la memoria para realizar el equilibrio de carga en un clúster. Como vSphere no es consciente de la agrupación en cluster ampliada, considera todos los hosts en ambos sitios al equilibrar la carga. Para evitar el tráfico entre sitios, NetApp recomienda configurar reglas de afinidad de DRS para gestionar una separación lógica de equipos virtuales. Esto garantizará que, a menos que se produzca un fallo completo del sitio, HA y DRS solo utilizarán los hosts locales.

Si crea una regla de afinidad de DRS para su clúster, puede especificar cómo aplica vSphere esa regla durante una conmutación al respaldo de una máquina virtual.

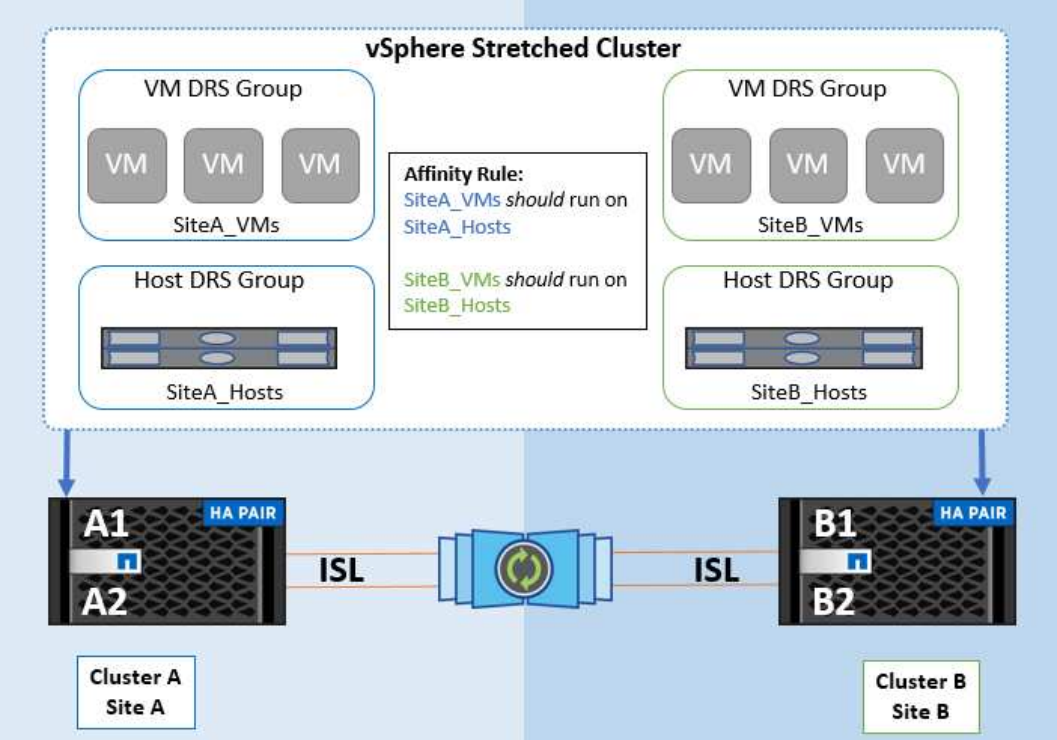
Hay dos tipos de reglas que se pueden especificar el comportamiento de conmutación al nodo de respaldo de alta disponibilidad de vSphere:

- Las reglas de anti-afinidad de equipos virtuales obligan a los equipos virtuales especificados a permanecer separados durante las acciones de recuperación tras fallos.
- Las reglas de afinidad de host de VM colocan las máquinas virtuales especificadas en un host particular o un miembro de un grupo definido de hosts durante las acciones de conmutación por error.

Mediante el uso de reglas de afinidad de host de VM en VMware DRS, se puede tener una separación lógica entre el sitio A y el sitio B, de modo que la VM se ejecute en el host en el mismo sitio que la cabina que está configurada como la controladora de lectura/escritura primaria para un almacén de datos determinado. Además, las reglas de afinidad de host de VM permiten que las máquinas virtuales permanezcan locales en el

almacenamiento, lo que, a su vez, verifica la conexión a la máquina virtual en caso de fallos de red entre los sitios.

A continuación se muestra un ejemplo de los grupos de hosts y las reglas de afinidad de las máquinas virtuales.



Best Practice

NetApp recomienda implementar reglas de «debería» en lugar de reglas de «debe» porque vSphere HA las infringe en caso de fallo. El uso de reglas «imprescindibles» podría provocar interrupciones del servicio.

La disponibilidad de los servicios debe prevalecer siempre sobre el rendimiento. En el caso en que falla un centro de datos completo, las reglas “must” deben elegir hosts del grupo de afinidad de host de VM y, cuando el centro de datos no esté disponible, las máquinas virtuales no se reiniciarán.

Implementación de VMware Storage DRS con NetApp MetroCluster

La función VMware Storage DRS permite agregar almacenes de datos en una sola unidad y equilibra los discos de máquinas virtuales cuando se superan los umbrales de control de I/O del almacenamiento.

El control de la I/O de almacenamiento se habilita de forma predeterminada en los clústeres DRS habilitados para Storage DRS. El control de las operaciones de I/O de almacenamiento permite a un administrador controlar la cantidad de I/O de almacenamiento que se asigna a máquinas virtuales durante periodos de congestión de I/O, lo que permite que las máquinas virtuales más importantes tengan preferencia por máquinas virtuales menos importantes para la asignación de recursos de E/S.

Storage DRS utiliza Storage vMotion para migrar los equipos virtuales a diferentes almacenes de datos dentro de un clúster de almacén de datos. En un entorno NetApp MetroCluster, una migración de máquinas virtuales debe controlarse dentro de los almacenes de datos de ese sitio. Por ejemplo, en condiciones ideales, la máquina virtual A, que se ejecuta en un host en el sitio A, debería migrar dentro de los almacenes de datos de la SVM en el sitio A. Si no lo hace, la máquina virtual continuará funcionando pero con un rendimiento degradado, ya que la lectura/escritura del disco virtual será desde la ubicación B a través de enlaces entre

sitios.

Best Practice

NetApp recomienda crear clústeres de almacenes de datos con respecto a la afinidad del sitio de almacenamiento; es decir, los almacenes de datos con afinidad del sitio A no se deben mezclar con clústeres de almacenes de datos con almacenes de datos con afinidad del sitio B.

Siempre que un equipo virtual se aprovisiona o se migra recientemente mediante Storage vMotion, NetApp recomienda actualizar manualmente todas las reglas de DRS de VMware específicas para dichos equipos virtuales. Esto determinará la afinidad de la máquina virtual en el nivel del sitio tanto para el host como para el almacén de datos y, por lo tanto, reducirá la sobrecarga de red y almacenamiento.

Directrices de implementación y diseño de VMSC

Este documento describe las guías de diseño e implementación para VMSC con sistemas de almacenamiento ONTAP.

Configuración de almacenamiento de NetApp

Las instrucciones de configuración para NetApp MetroCluster (en lo que se refiere como configuración de MCC) están disponibles en "Documentación de MetroCluster". También puede encontrar instrucciones para SnapMirror Active Sync en "Información general sobre la continuidad del negocio de SnapMirror".

Después de configurar MetroCluster, administrarlo es como administrar un entorno ONTAP tradicional. Puede configurar máquinas virtuales de almacenamiento (SVM) con diferentes herramientas, como la interfaz de línea de comandos (CLI), System Manager o Ansible. Una vez que se han configurado las SVM, cree interfaces lógicas (LIF), volúmenes y números de unidad lógica (LUN) en el clúster que se utilizarán para operaciones normales. Estos objetos se replicarán automáticamente en el otro clúster mediante la red de conexión de clústeres.

Si no utiliza MetroCluster, puede usar la sincronización activa de SnapMirror, que proporciona protección granular de almacenes de datos y acceso activo-activo en múltiples clústeres de ONTAP en diferentes dominios de fallo. SnapMirror Active Sync utiliza grupos de coherencia para garantizar la coherencia en orden de escritura entre uno o varios almacenes de datos y puede crear varios grupos de coherencia en función de los requisitos de la aplicación y del almacén de datos. Los grupos de coherencia son especialmente útiles para aplicaciones que requieren sincronización de datos entre varios almacenes de datos. La sincronización activa de SnapMirror también admite asignaciones de dispositivos sin formato (RDM) y almacenamiento conectado mediante invitado con iniciadores iSCSI invitados. Puede obtener más información sobre grupos de consistencia en "Información general sobre los grupos de consistencia".

Hay alguna diferencia en la gestión de una configuración VMSC con sincronización activa de SnapMirror en comparación con una MetroCluster. En primer lugar, se trata de una configuración solo SAN, no se puede proteger ningún almacén de datos NFS con sincronización activa de SnapMirror. Segundo, debe asignar ambas copias de las LUN a los hosts ESXi para que accedan a los almacenes de datos replicados en ambos dominios de fallo.

VMware vSphere ha

Cree un clúster de vSphere HA

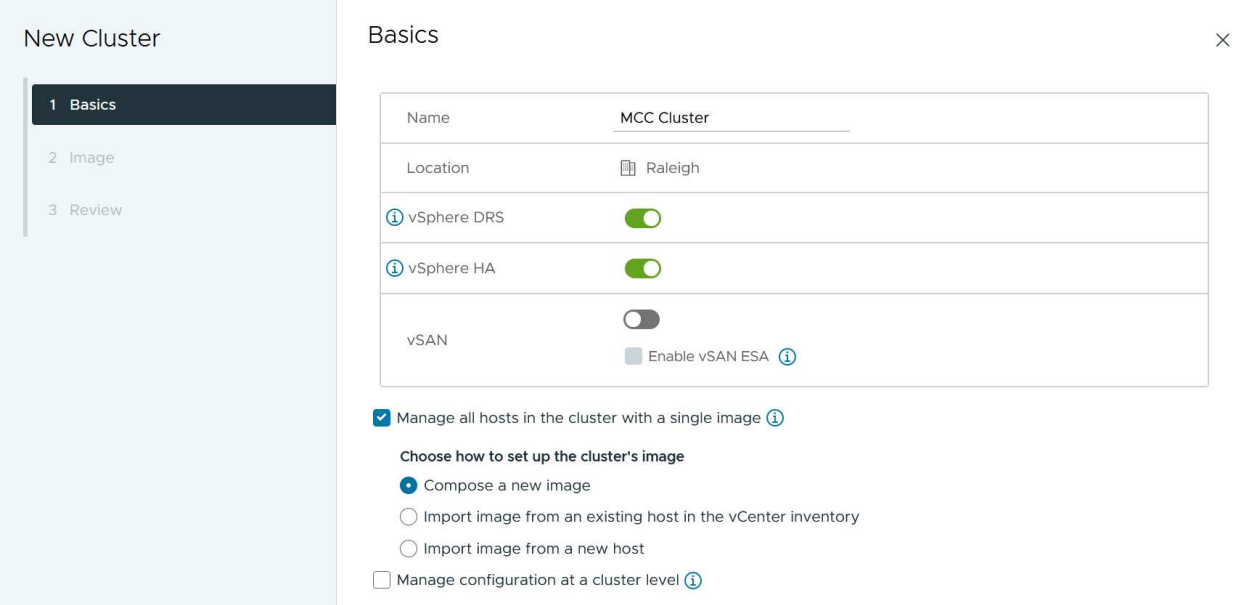
La creación de un clúster de vSphere HA es un proceso de varios pasos que se documenta completamente en "Cómo se crean y configuran clústeres en vSphere Client en docs.vmware.com". En resumen, primero debe

crear un clúster vacío y, después, utilizando vCenter, debe añadir hosts y especificar la alta disponibilidad de vSphere y otros ajustes del clúster.

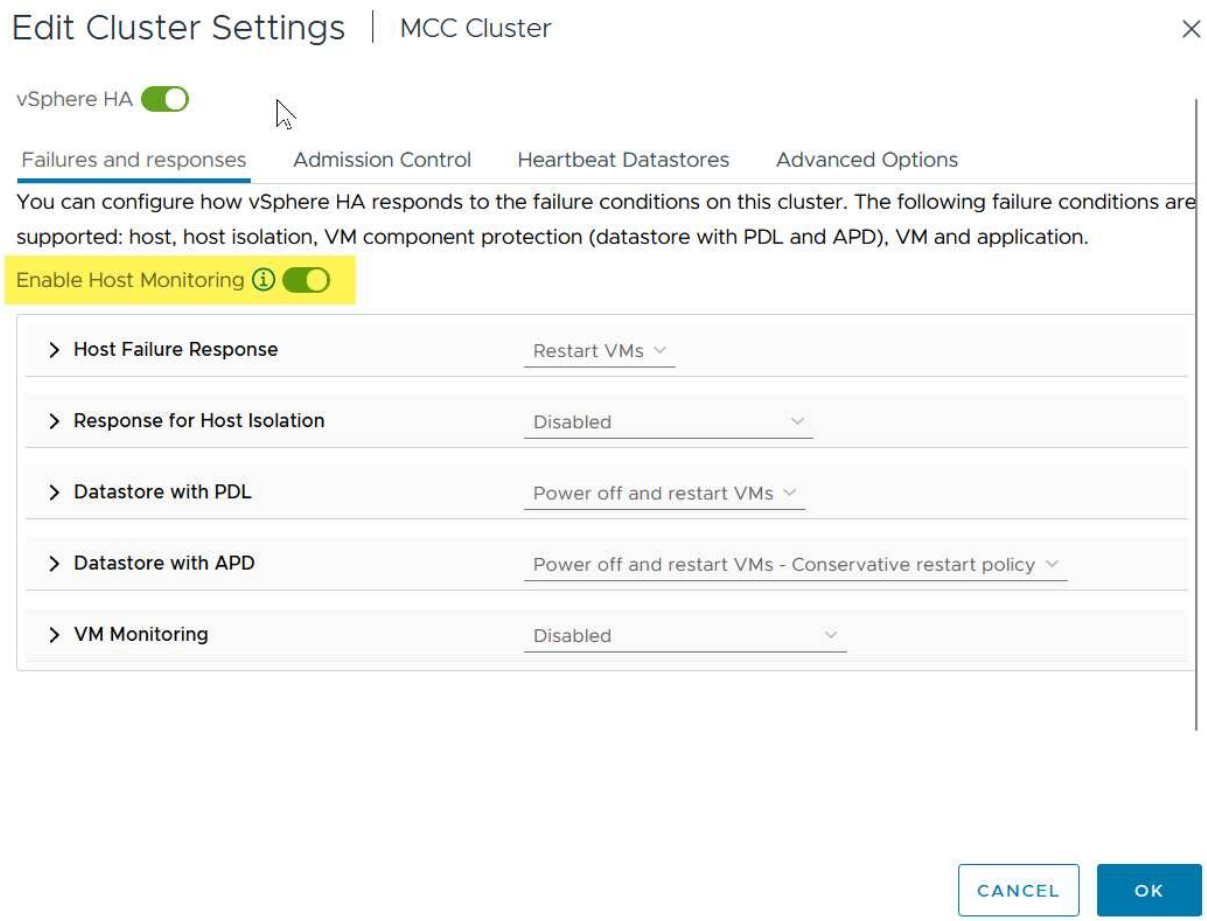
Nota: Nada en este documento reemplaza "Prácticas recomendadas para VMware vSphere Metro Storage Cluster"

Para configurar un clúster de alta disponibilidad, realice los siguientes pasos:

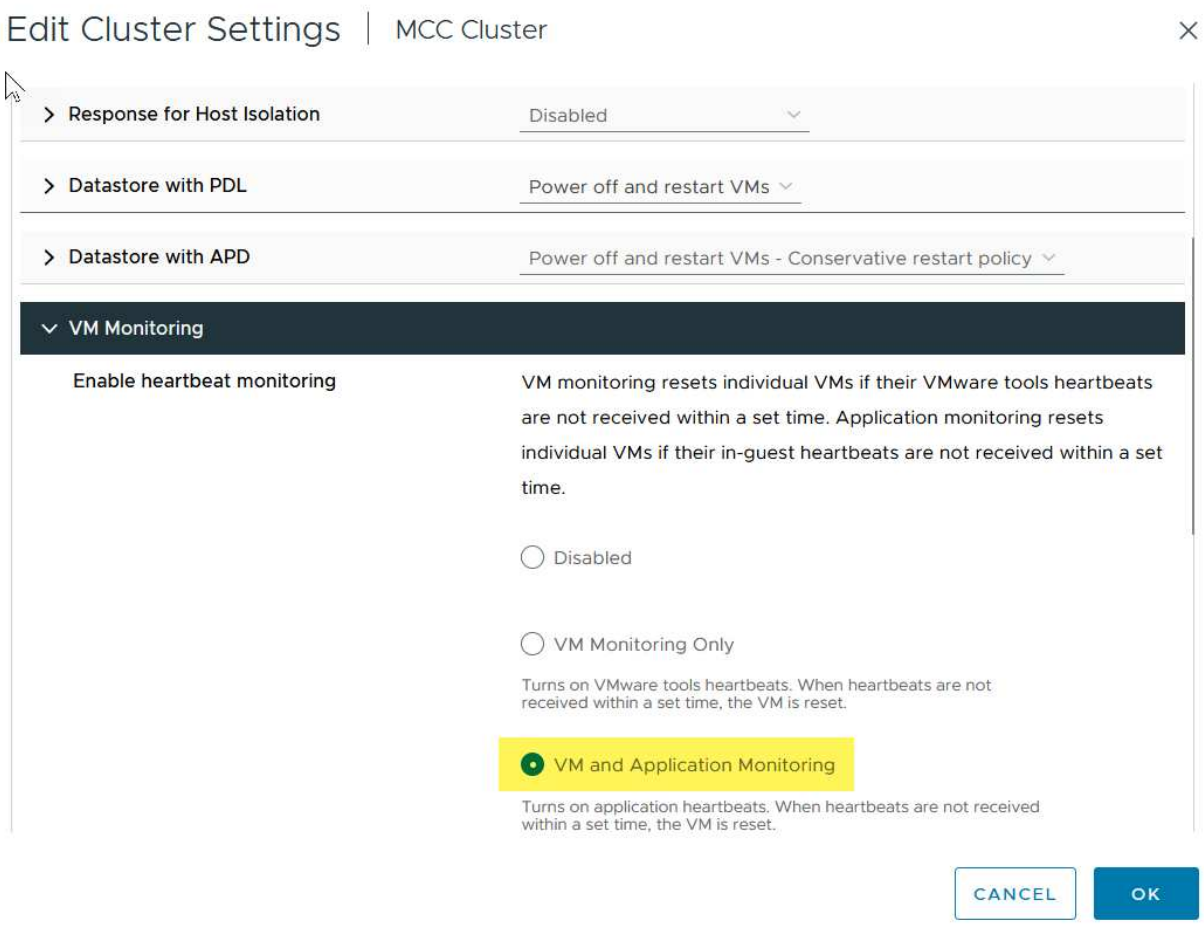
1. Conéctese a la interfaz de usuario de vCenter.
2. En Hosts and Clusters, vaya al centro de datos donde desea crear su clúster de alta disponibilidad.
3. Haga clic con el botón derecho en el objeto del centro de datos y seleccione New Cluster. En los conceptos básicos, asegúrese de haber habilitado vSphere DRS y vSphere HA. Complete el asistente.



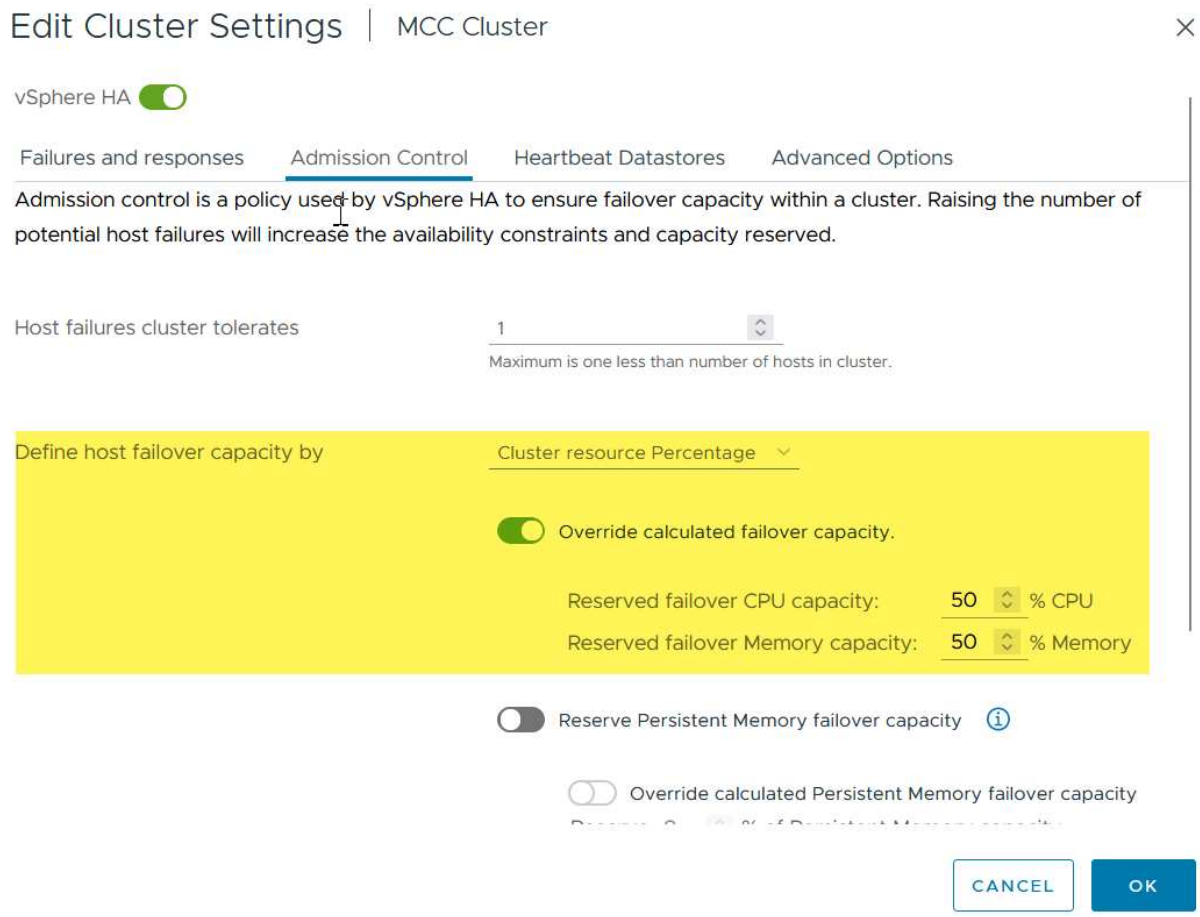
1. Seleccione el clúster y vaya a la pestaña Configure. Seleccione vSphere HA y haga clic en Edit.
2. En Supervisión de host, seleccione la opción Habilitar supervisión de host.



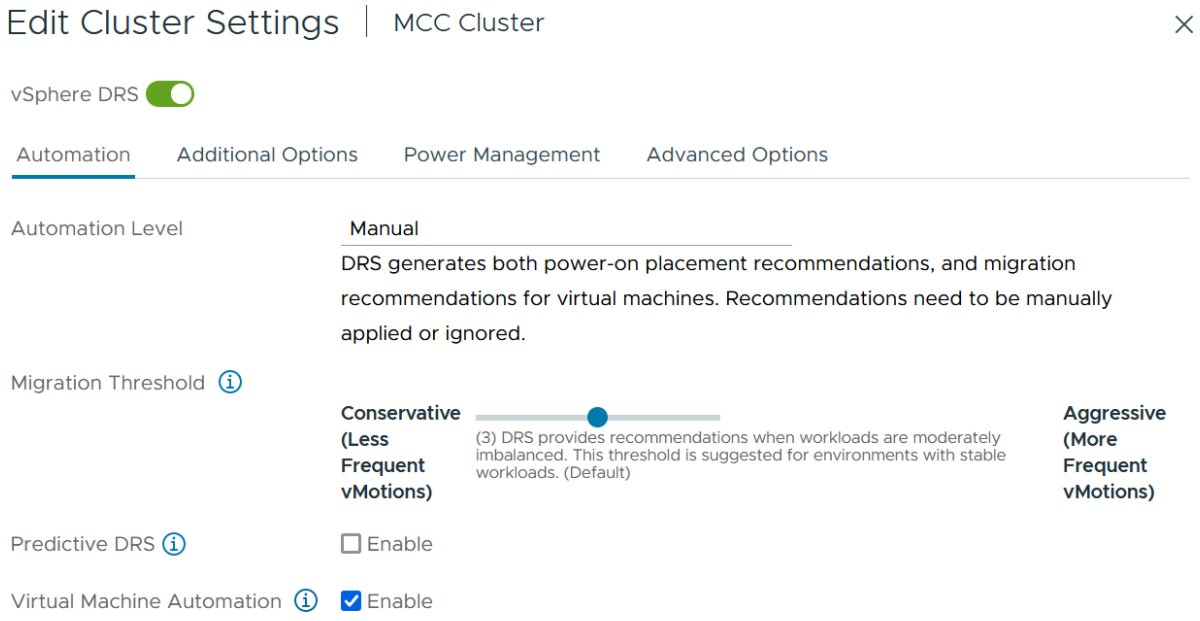
1. Mientras todavía está en la pestaña Fallos y Respuestas, en VM Monitoring, seleccione la opción VM Monitoring Only o VM and Application Monitoring.



1. En Control de admisión, establezca la opción de control de admisión de HA en Reserva de recursos de cluster; utilice 50% CPU/MEM.



1. Se hace clic en «OK».
2. Seleccione DRS y haga clic en EDIT.
3. Establezca el nivel de automatización en manual a menos que las aplicaciones lo requieran.



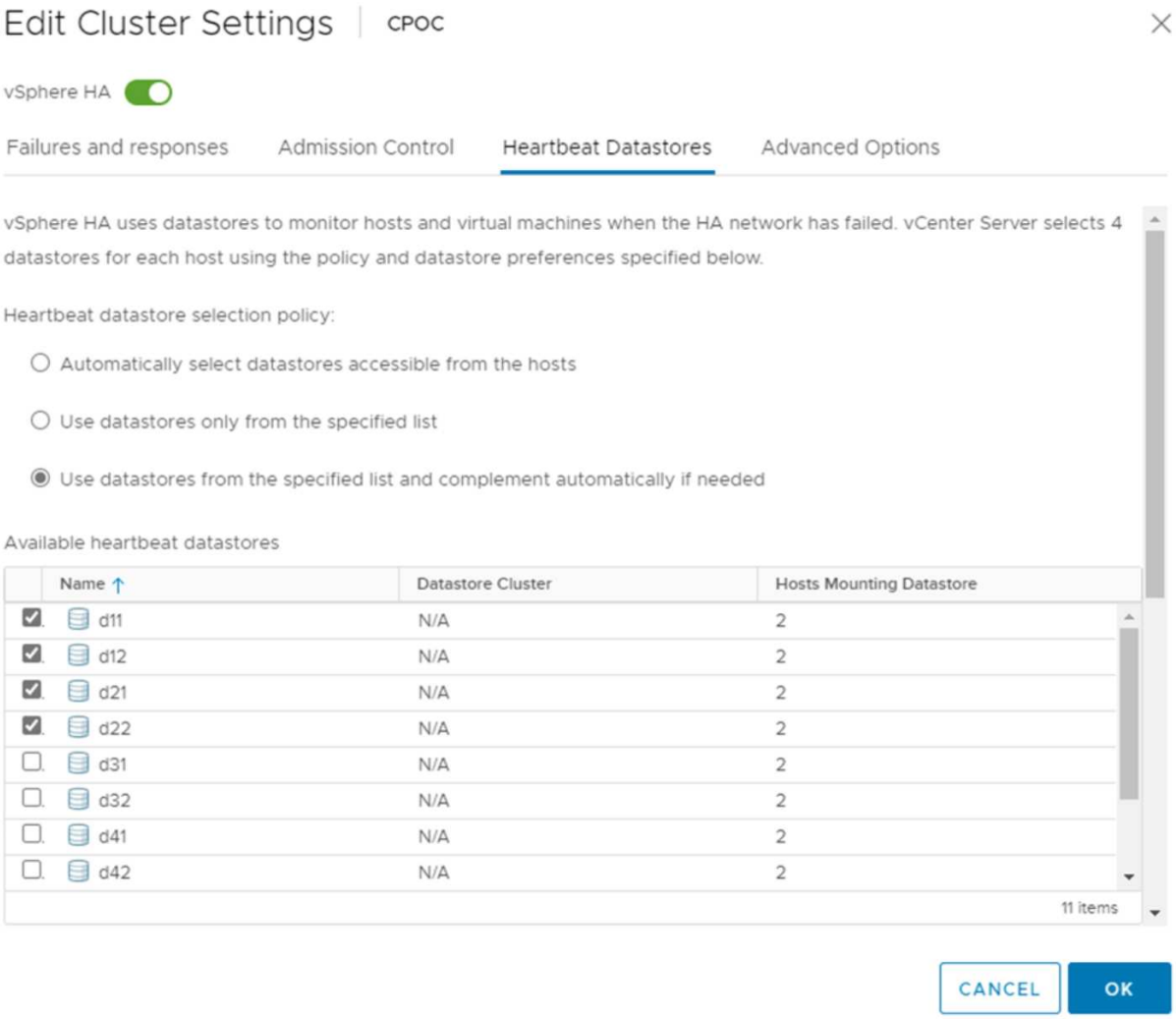
1. Habilite VM Component Protection, consulte "docs.vmware.com".
2. Se recomiendan las siguientes configuraciones adicionales de alta disponibilidad de vSphere para VMSC con MCC:

Fallo	Respuesta
Error del host	Reiniciar las máquinas virtuales
Aislamiento de hosts	Deshabilitado
Almacén de datos con pérdida permanente de dispositivo (PDL)	Apagar y reiniciar los equipos virtuales
Almacén de datos con todas las rutas inactivas (APD)	Apagar y reiniciar los equipos virtuales
El huésped no es molesto	Restablecer las máquinas virtuales
Política de reinicio de máquinas virtuales	Determinado por la importancia del equipo virtual
Respuesta para el aislamiento del host	Apagar y reiniciar equipos virtuales
Respuesta para datastore con PDL	Apagar y reiniciar los equipos virtuales
Respuesta del almacén de datos con APD	Apagar y reiniciar equipos virtuales (conservador)
Demora en recuperación tras fallos de equipos virtuales para APD	3 minutos
Respuesta para la recuperación de APD con tiempo de espera APD	Deshabilitado
Supervisión de la sensibilidad de los equipos virtuales	Preajuste ALTO

Configurar almacenes de datos para Heartbeat

La alta disponibilidad de vSphere utiliza almacenes de datos para supervisar hosts y máquinas virtuales cuando se produce un error en la red de gestión. Es posible configurar la forma en la que vCenter selecciona los almacenes de datos de latido. Para configurar los almacenes de datos para latir, lleve a cabo los siguientes pasos:

1. En la sección Datastore Heartbeat, seleccione Use datastores from the Specified List y complemente automáticamente si es necesario.
2. Seleccione los almacenes de datos que desee utilizar vCenter en ambos sitios y pulse OK.

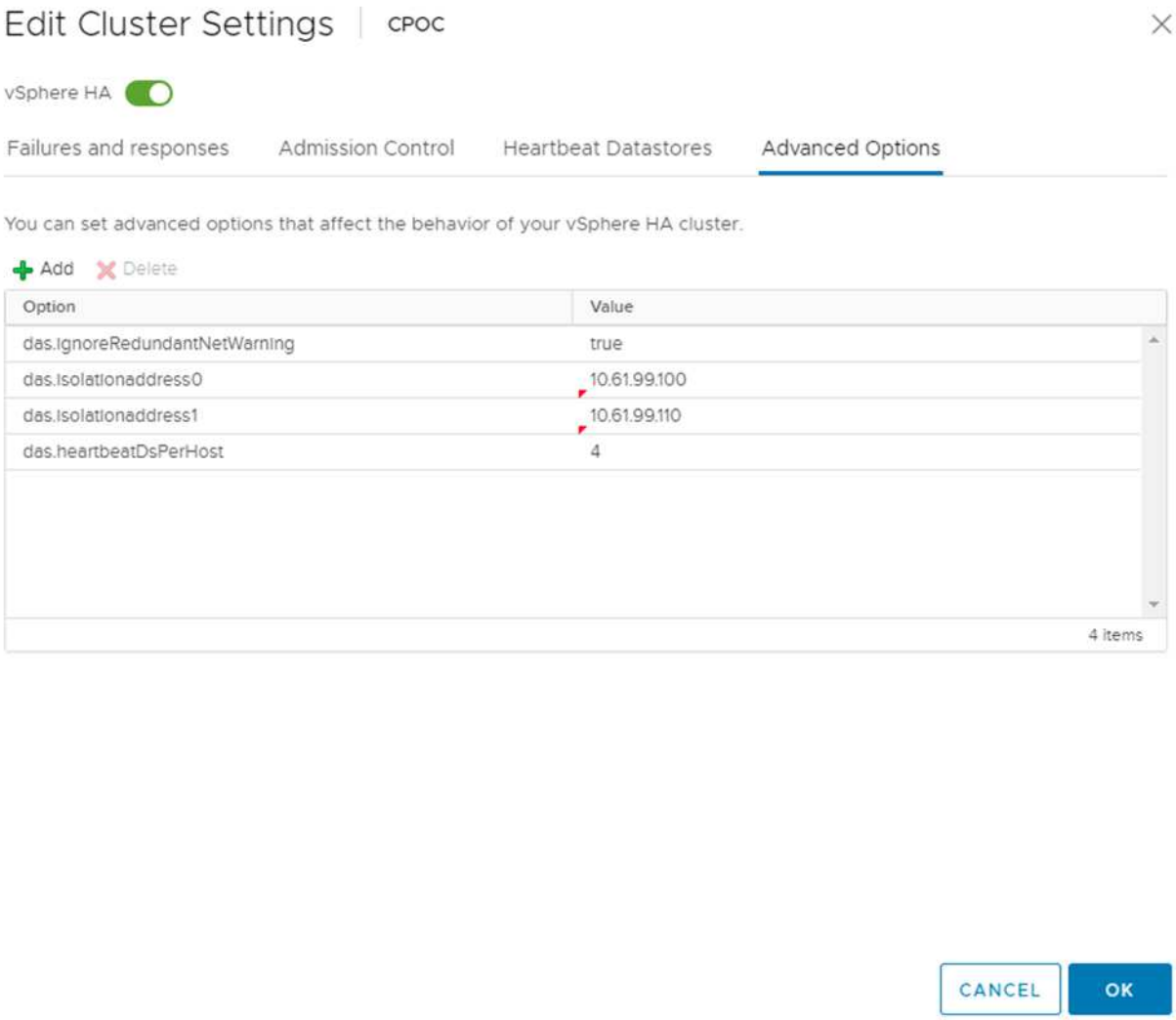


Configurar opciones avanzadas

- Detección de fallos del host *

Los eventos de aislamiento se producen cuando los hosts dentro de un clúster de alta disponibilidad pierden la conectividad a la red u otros hosts del clúster. De forma predeterminada, vSphere HA utilizará la puerta de enlace predeterminada para su red de gestión como dirección de aislamiento predeterminada. Sin embargo, puede especificar direcciones de aislamiento adicionales para que el host haga ping para determinar si se debe activar una respuesta de aislamiento. Agregue dos IP de aislamiento que puedan hacer ping, una por sitio. No utilice la IP de la puerta de enlace. La configuración avanzada de HA de vSphere utilizada es das.isolationaddress. Puede utilizar las direcciones IP de ONTAP o Mediator para este fin.

Consulte "core.vmware.com" para obtener más información.



Agregar una configuración avanzada llamada das.heartbeatDsPerHost puede aumentar el número de almacenes de datos de latido. Utilice cuatro almacenes de datos para el corazón (HB DSS): Dos por sitio. Utilice la opción “Seleccionar de la lista pero cumplido”. Esto es necesario porque si un sitio falla, usted todavía necesita dos HB DSS. Sin embargo, esas empresas no tienen que estar protegidas con sincronización activa de SnapMirror o MCC.

Consulte "core.vmware.com" para obtener más información.

Afinidad de VMware DRS para NetApp MetroCluster

En esta sección creamos grupos DRS para equipos virtuales y hosts para cada sitio\clúster del entorno MetroCluster. A continuación, configuramos las reglas de VM\Host para alinear la afinidad de host de VM con los recursos de almacenamiento local. Por ejemplo, las máquinas virtuales de la dirección A pertenecen al grupo de máquinas virtuales sitea_vms y la ubicación A pertenecen al grupo de hosts sitea_hosts. A continuación, en VM\Host Rules, indicamos que sitea_vms debe ejecutarse en hosts en sitea_Hosts.

Best Practice

- NetApp recomienda encarecidamente la especificación **Debe ejecutarse en hosts del grupo** en lugar de la especificación **Debe ejecutarse en hosts del grupo**. En caso de que se produzca un fallo del host del sitio A, es necesario reiniciar las máquinas virtuales del sitio A en los hosts del sitio B a través de vSphere HA, pero la última especificación no permite a HA reiniciar los equipos virtuales en el sitio B, ya que es una

regla estricta. La especificación anterior es una regla flexible y se infringirá en caso de alta disponibilidad, lo que permitirá la disponibilidad en lugar de rendimiento.

Nota: Puede crear una alarma basada en eventos que se activa cuando una máquina virtual viola una regla de afinidad VM-Host. En vSphere Client, agregue una nueva alarma para la máquina virtual y seleccione “VM is Violating VM-Host Affinity Rule” como disparador de eventos. Para obtener más información sobre la creación y edición de alarmas, consulte "Supervisión y rendimiento de vSphere" documentación.

Crear grupos de hosts DRS

Para crear grupos de hosts DRS específicos del sitio A y del sitio B, realice los siguientes pasos:

1. En vSphere Web Client, haga clic con el botón derecho en el clúster en el inventario y seleccione Settings.
2. Haga clic en VM\Host Groups.
3. Haga clic en Añadir.
4. Escriba el nombre del grupo (por ejemplo, sitea_hosts).
5. En el menú Tipo, seleccione Grupo de hosts.
6. Haga clic en Agregar y seleccione los hosts deseados del sitio A y haga clic en Aceptar.
7. Repita estos pasos para agregar otro grupo de hosts para el sitio B.
8. Haga clic en Aceptar.

Crear grupos de máquinas virtuales DRS

Para crear grupos de máquinas virtuales DRS específicos del sitio A y del sitio B, realice los siguientes pasos:

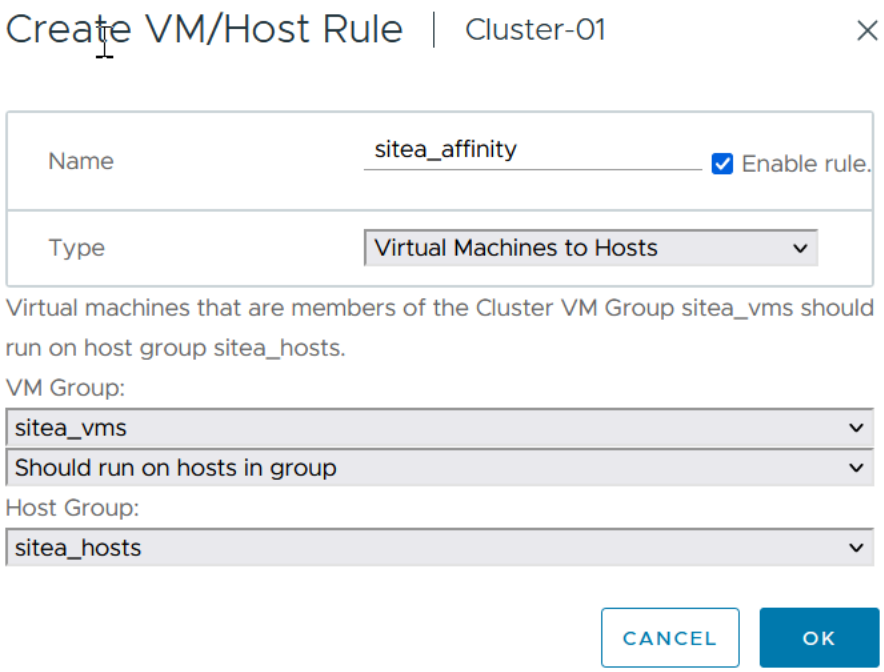
1. En vSphere Web Client, haga clic con el botón derecho en el clúster en el inventario y seleccione Settings.
2. Haga clic en VM\Host Groups.
3. Haga clic en Añadir.
4. Escriba el nombre del grupo (por ejemplo, sitea_vms).
5. En el menú Type, seleccione VM Group.
6. Haga clic en Add y seleccione las máquinas virtuales deseadas en el sitio A y, a continuación, haga clic en OK.
7. Repita estos pasos para agregar otro grupo de hosts para el sitio B.
8. Haga clic en Aceptar.

Crear reglas de host de VM

Para crear reglas de afinidad de DRS específicas para el sitio A y el sitio B, realice los siguientes pasos:

1. En vSphere Web Client, haga clic con el botón derecho en el clúster en el inventario y seleccione Settings.
2. Haga clic en VM\Host Rules.
3. Haga clic en Añadir.
4. Escriba el nombre de la regla (por ejemplo, sitea_affinity).
5. Compruebe que la opción Activar regla está activada.
6. En el menú Type, seleccione Virtual Machines to Hosts.

7. Seleccione el grupo de VM (por ejemplo, sitea_vms).
8. Seleccione el grupo Host (por ejemplo, sitea_Hosts).
9. Repita estos pasos para añadir otra regla VM\Host para el sitio B.
10. Haga clic en Aceptar.



DRS de almacenamiento de VMware vSphere para NetApp MetroCluster

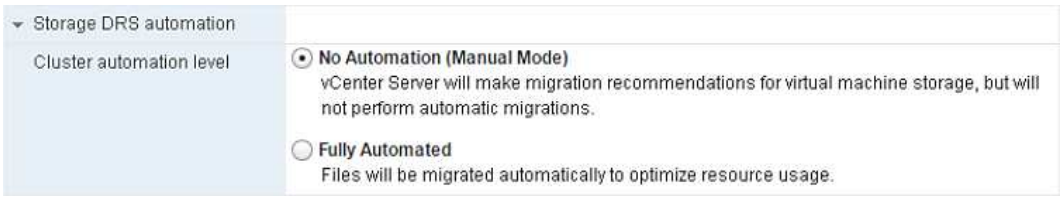
Crear clústeres de almacenes de datos

Para configurar un clúster de almacén de datos para cada sitio, complete los siguientes pasos:

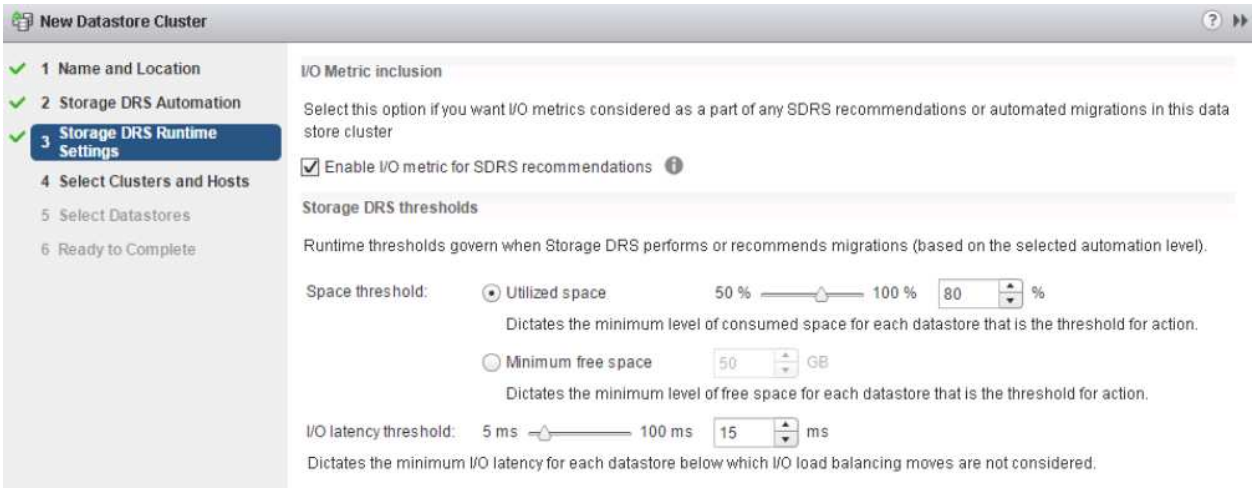
1. Use el cliente web de vSphere, vaya al centro de datos donde reside el clúster de alta disponibilidad en Storage.
2. Haga clic con el botón derecho en el objeto del centro de datos y seleccione Storage > New Datastore Cluster.
3. Seleccione la opción ON Storage DRS y haga clic en Next.
4. Establezca todas las opciones en Sin automatización (Modo manual) y haga clic en Siguiente.

Best Practice

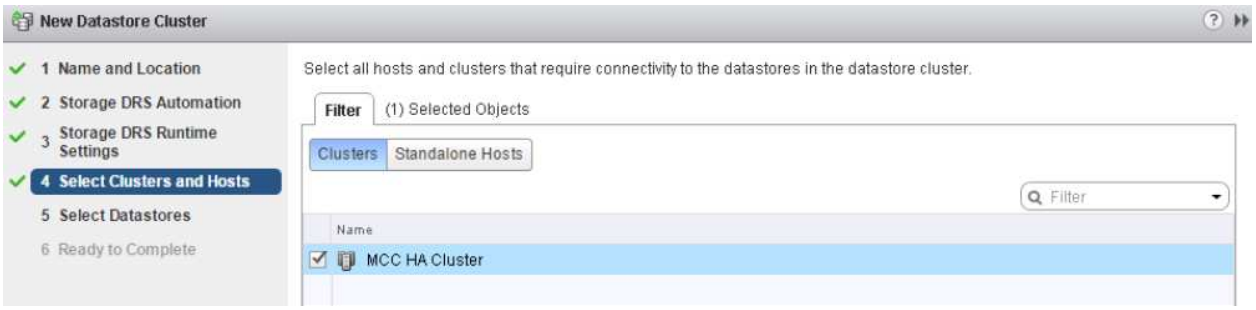
- NetApp recomienda configurar el DRS de almacenamiento en modo manual, de modo que el administrador decida y controle cuándo es necesario realizar las migraciones.



1. Compruebe que la casilla de verificación Activar Métrica de E/S para Recomendaciones de SDRS está activada; los valores de métrica se pueden dejar con los valores predeterminados.



1. Seleccione el clúster de alta disponibilidad y haga clic en Next.



1. Seleccione los almacenes de datos que pertenecen al sitio A y haga clic en Next.

1. Revise las opciones y haga clic en Finish.
2. Repita estos pasos para crear el clúster de almacenes de datos del sitio B y verifique que solo estén seleccionados los almacenes de datos del sitio B.

Disponibilidad del vCenter Server

Los dispositivos vCenter Server Appliances (VCSA) deben estar protegidos con alta disponibilidad de vCenter. La alta disponibilidad de vCenter le permite implementar dos VCSA en un par de alta disponibilidad activo-pasivo. Uno en cada dominio de fallo. Puede obtener más información sobre la alta disponibilidad de vCenter en "docs.vmware.com".

Resiliencia para eventos planificados y no planificados

NetApp MetroCluster y SnapMirror Active Sync son potentes herramientas que mejoran la alta disponibilidad y las operaciones no disruptivas del hardware de NetApp y del software ONTAP®.

Estas herramientas proporcionan protección en todo el sitio para todo el entorno de almacenamiento, lo que garantiza que los datos están siempre disponibles. Ya sea que utilice servidores independientes, clústeres de servidores de alta disponibilidad, contenedores Docker o servidores virtualizados, la tecnología NetApp mantiene fácilmente la disponibilidad de almacenamiento en caso de interrupción total por pérdida de alimentación, refrigeración o conectividad de red, apagado del array de almacenamiento o error de funcionamiento.

MetroCluster y SnapMirror de sincronización activa proporcionan tres métodos básicos de continuidad de datos en caso de eventos previstos o no planificados:

- Componentes redundantes para protección contra fallos de un solo componente
- Toma de control local de alta disponibilidad para eventos que afectan a una única controladora
- Protección completa del sitio: Reanudación rápida del servicio al mover el almacenamiento y el acceso de clientes del clúster de origen al clúster de destino

Esto significa que las operaciones continúan sin problemas en caso de fallo de un único componente y vuelven automáticamente al funcionamiento redundante cuando se reemplaza el componente fallido.

Todos los clústeres de ONTAP, excepto los clústeres de un solo nodo (normalmente las versiones definidas por software, como ONTAP Select, por ejemplo), tienen funciones de alta disponibilidad incorporadas denominadas toma de control y retorno al nodo primario. Cada controladora del clúster se empareja con otra controladora, lo que forma una pareja de alta disponibilidad. Estos pares garantizan que cada nodo esté conectado localmente al almacenamiento.

La toma de control es un proceso automatizado en el que un nodo asume el almacenamiento del otro para mantener los servicios de datos. Giveback es el proceso inverso que restaura el funcionamiento normal. La toma de control puede planificarse, por ejemplo, al realizar tareas de mantenimiento del hardware o actualizaciones de ONTAP, o no planificadas, resultantes de un error de hardware o de alarma en el nodo.

Durante una toma de control, las interfaces lógicas de almacenamiento conectadas a red (LIF NAS) en configuraciones de MetroCluster conmutan automáticamente al respaldo. Sin embargo, los LIF de red de área de almacenamiento (LIF SAN) no conmutan al nodo de respaldo; seguirán utilizando la ruta directa a los números de unidad lógica (LUN).

Si quiere más información sobre la toma de control y el retorno al nodo primario de alta disponibilidad, consulte la "Información general sobre la gestión de parejas de HA". Vale la pena señalar que esta funcionalidad no es específica de la sincronización activa de MetroCluster o SnapMirror.

El cambio de sitio con MetroCluster se produce cuando un sitio está sin conexión o como una actividad planificada para el mantenimiento de todo el sitio. El sitio restante asume la propiedad de los recursos de almacenamiento (discos y agregados) del clúster sin conexión y las SVM del sitio con el que se ha producido el fallo se conectan y se reinician en el sitio de desastre, conservando su identidad completa para el acceso de clientes y host.

Con la sincronización activa de SnapMirror, dado que ambas copias se usan de forma activa a la vez, los hosts existentes seguirán funcionando. El Mediador de NetApp es necesario para garantizar que la conmutación por error del sitio se produce correctamente.

Escenarios de fallo para VMSC con MCC

En las siguientes secciones se resumen los resultados esperados de varios escenarios de fallo con sistemas VMSC y NetApp MetroCluster.

Fallo de ruta de almacenamiento única

En esta situación, si se produce un error en componentes como el puerto HBA, el puerto de red, el puerto del switch de datos de interfaz de usuario o un cable FC o Ethernet, esa ruta particular al dispositivo de almacenamiento se marca como muerta por el host ESXi. Si se configuran varias rutas para el dispositivo de almacenamiento proporcionando resiliencia en el puerto de HBA/red/switch, ESXi idealmente ejecuta una conmutación de rutas. Durante este periodo, las máquinas virtuales permanecen en ejecución sin que se vean afectadas, porque se cuida de la disponibilidad del almacenamiento mediante varias rutas al dispositivo de almacenamiento.

Nota: No hay cambios en el comportamiento de MetroCluster en este escenario, y todos los almacenes de datos siguen intactos desde sus respectivos sitios.

Best Practice

En entornos en los que se utilizan volúmenes NFS/iSCSI, NetApp recomienda tener al menos dos vínculos superiores de red configurados para el puerto NFS vmkernel en el vSwitch estándar y lo mismo en el grupo de puertos en el que se asigna la interfaz de NFS vmkernel para el vSwitch distribuido. La agrupación de NIC se puede configurar en activo-activo o activo-en espera.

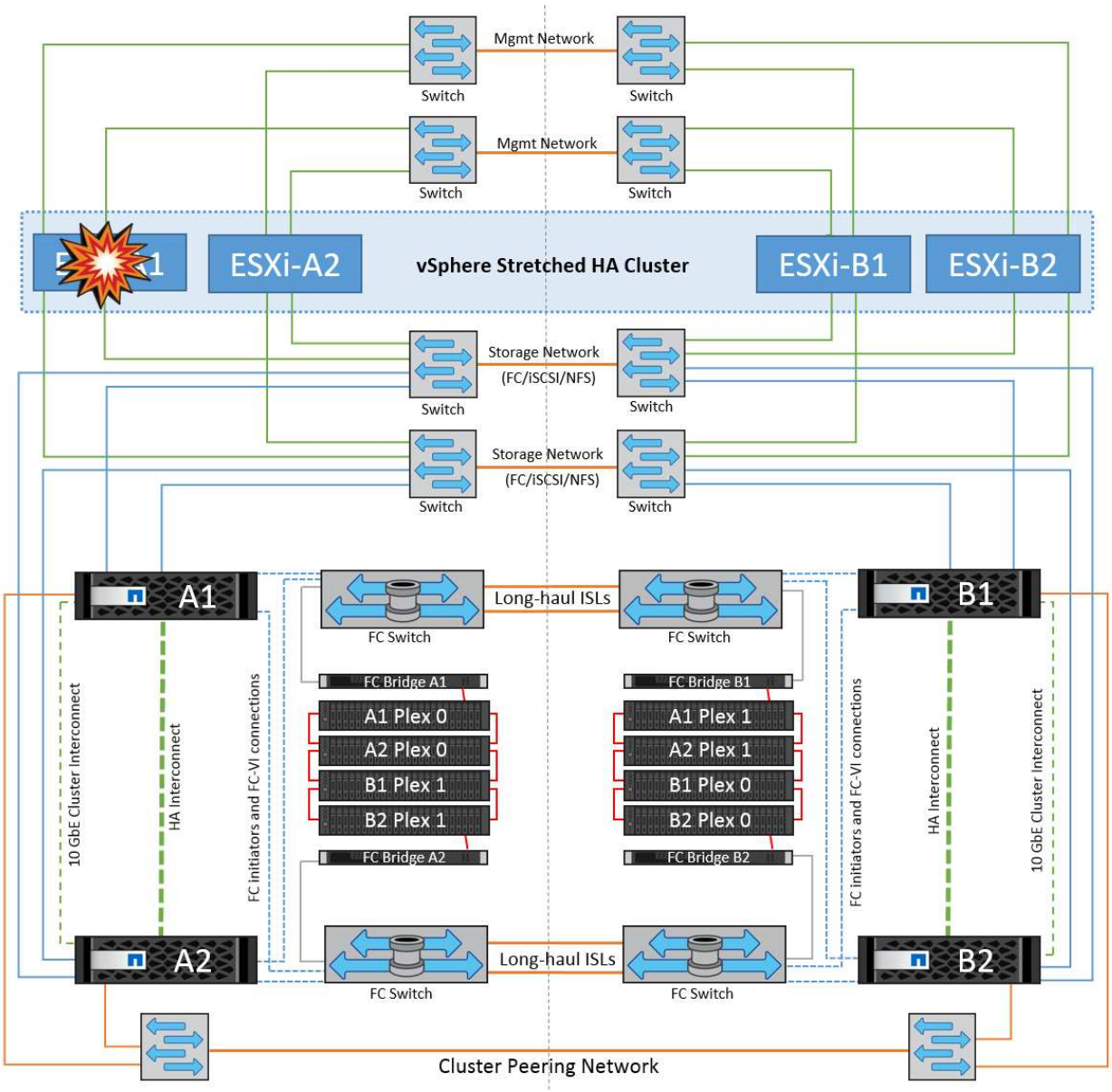
Además, para las LUN iSCSI, la multivía debe configurarse vinculando las interfaces de vmkernel con los adaptadores de red iSCSI. Si quiere más información, consulte la documentación de almacenamiento de vSphere.

Best Practice

En entornos en los que se usan LUN de Fibre Channel, NetApp recomienda tener al menos dos HBA, lo que garantiza la resistencia a nivel de HBA/puerto. NetApp también recomienda la división en zonas de un solo iniciador a un único destino como práctica recomendada para configurar la división en zonas.

Debe utilizarse Virtual Storage Console (VSC) para establecer normativas de accesos múltiples, porque establece normativas para todos los dispositivos de almacenamiento de NetApp nuevos y existentes.

Fallo de un host ESXi único



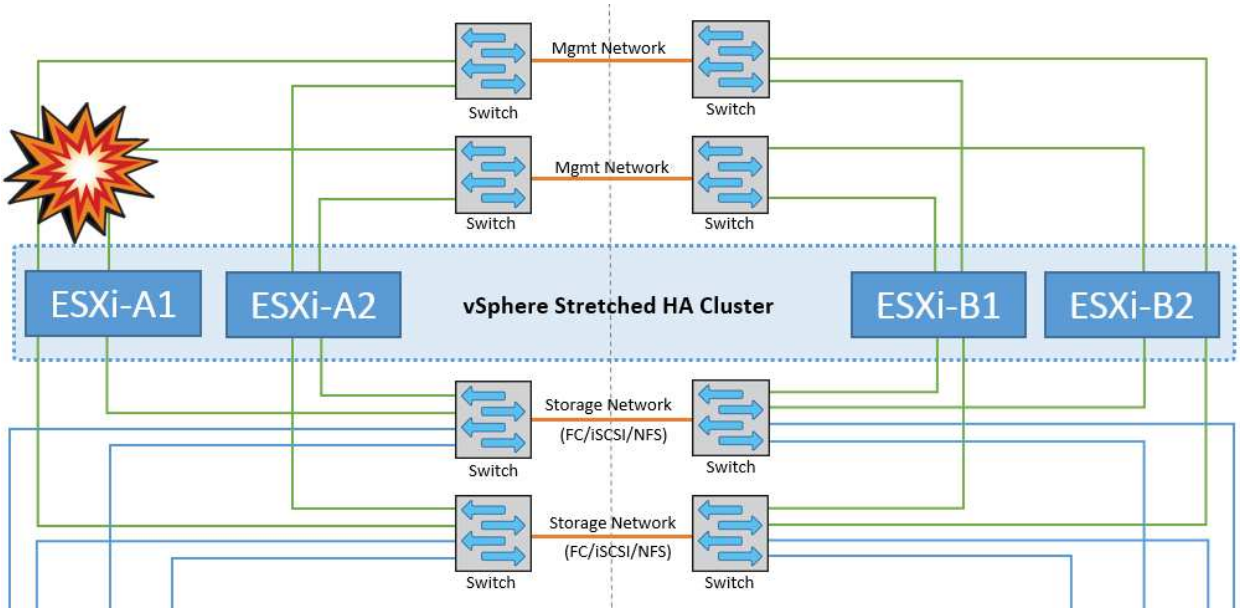
En esta situación, si hay un fallo de host ESXi, el nodo maestro del clúster de alta disponibilidad de VMware detecta el fallo del host porque ya no recibe los latidos de red. Para determinar si el host está realmente inactivo o sólo una partición de red, el nodo maestro supervisa los latidos del almacén de datos y, si están ausentes, realiza una comprobación final haciendo ping en las direcciones IP de gestión del host fallido. Si todas estas comprobaciones son negativas, el nodo maestro declara a este host un host fallido y todas las máquinas virtuales que se estaban ejecutando en este host fallido se reinician en el host superviviente del cluster.

Si se han configurado las reglas de afinidad de host y VM de DRS (las VM del grupo de VM sitea_vms deben ejecutar hosts en el grupo de hosts sitea_Hosts), el maestro de HA primero comprueba los recursos disponibles en el sitio A. Si no hay hosts disponibles en el sitio A, el maestro intenta reiniciar las máquinas virtuales en los hosts del sitio B.

Es posible que las máquinas virtuales se inicien en los hosts ESXi en el otro sitio si hay una restricción de recursos en el sitio local. Sin embargo, las reglas de afinidad de host y máquina virtual de DRS definidas corregirán si se viola alguna regla migrando las máquinas virtuales de nuevo a cualquier host ESXi sobreviviente en el sitio local. En los casos en que DRS se defina en manual, NetApp recomienda invocar DRS y aplicar las recomendaciones para corregir la ubicación de la máquina virtual.

No hay ningún cambio en el comportamiento de MetroCluster en este escenario y todos los almacenes de datos siguen estando intactos en sus sitios respectivos.

Aislamiento de hosts ESXi

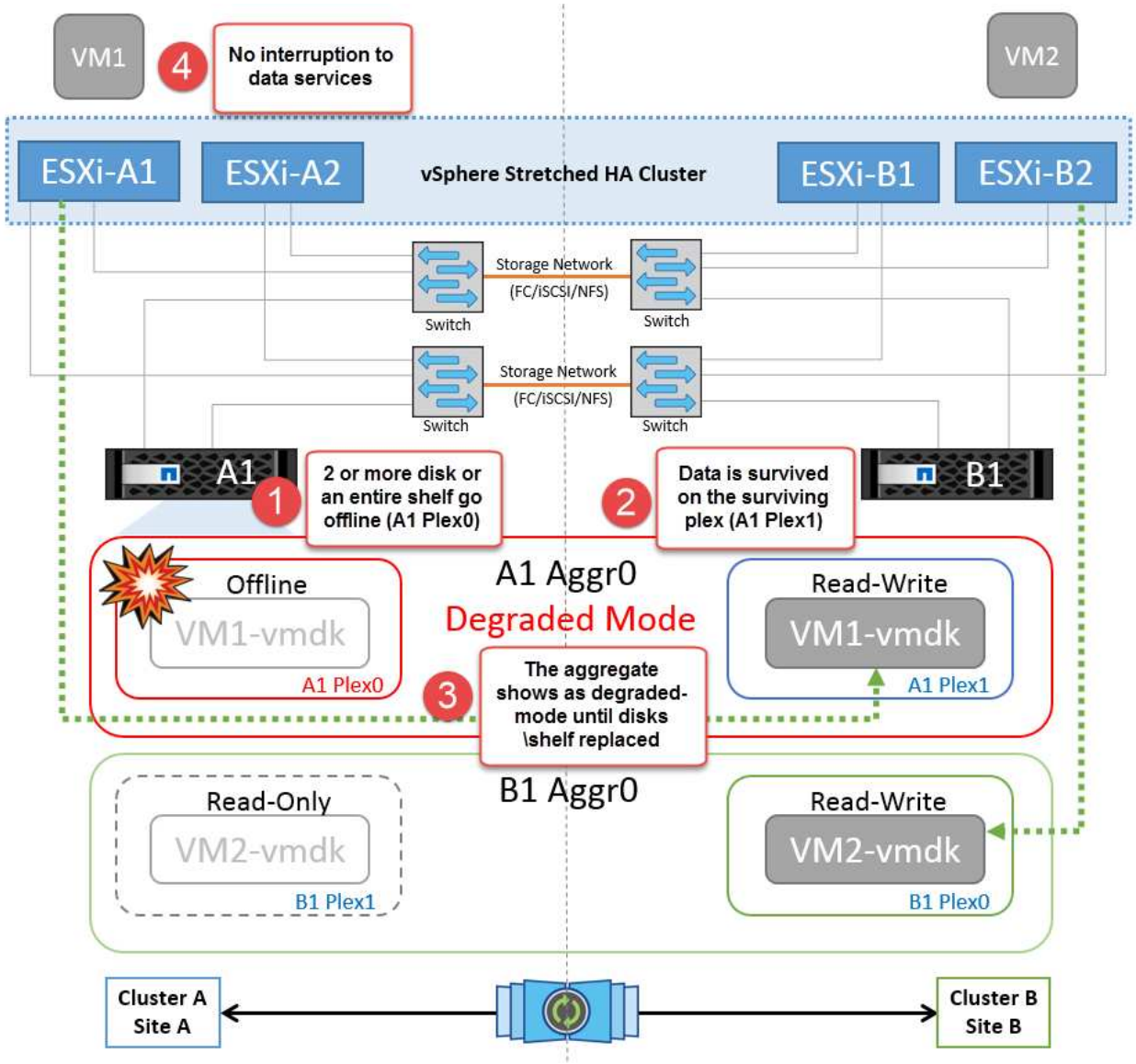


En esta situación, si la red de gestión del host ESXi está inactiva, el nodo principal del clúster de alta disponibilidad no recibirá ningún latido y, por lo tanto, este host se aísla en la red. Para determinar si ha fallado o solo está aislado, el nodo maestro comienza a supervisar el latido del almacén de datos. Si está presente, el nodo maestro declara que el host está aislado. Dependiendo de la respuesta de aislamiento configurada, el host puede optar por apagarse, apagar las máquinas virtuales o incluso dejar encendidas las máquinas virtuales. El intervalo predeterminado para la respuesta de aislamiento es de 30 segundos.

No hay ningún cambio en el comportamiento de MetroCluster en este escenario y todos los almacenes de datos siguen estando intactos en sus sitios respectivos.

Fallo de la bandeja de discos

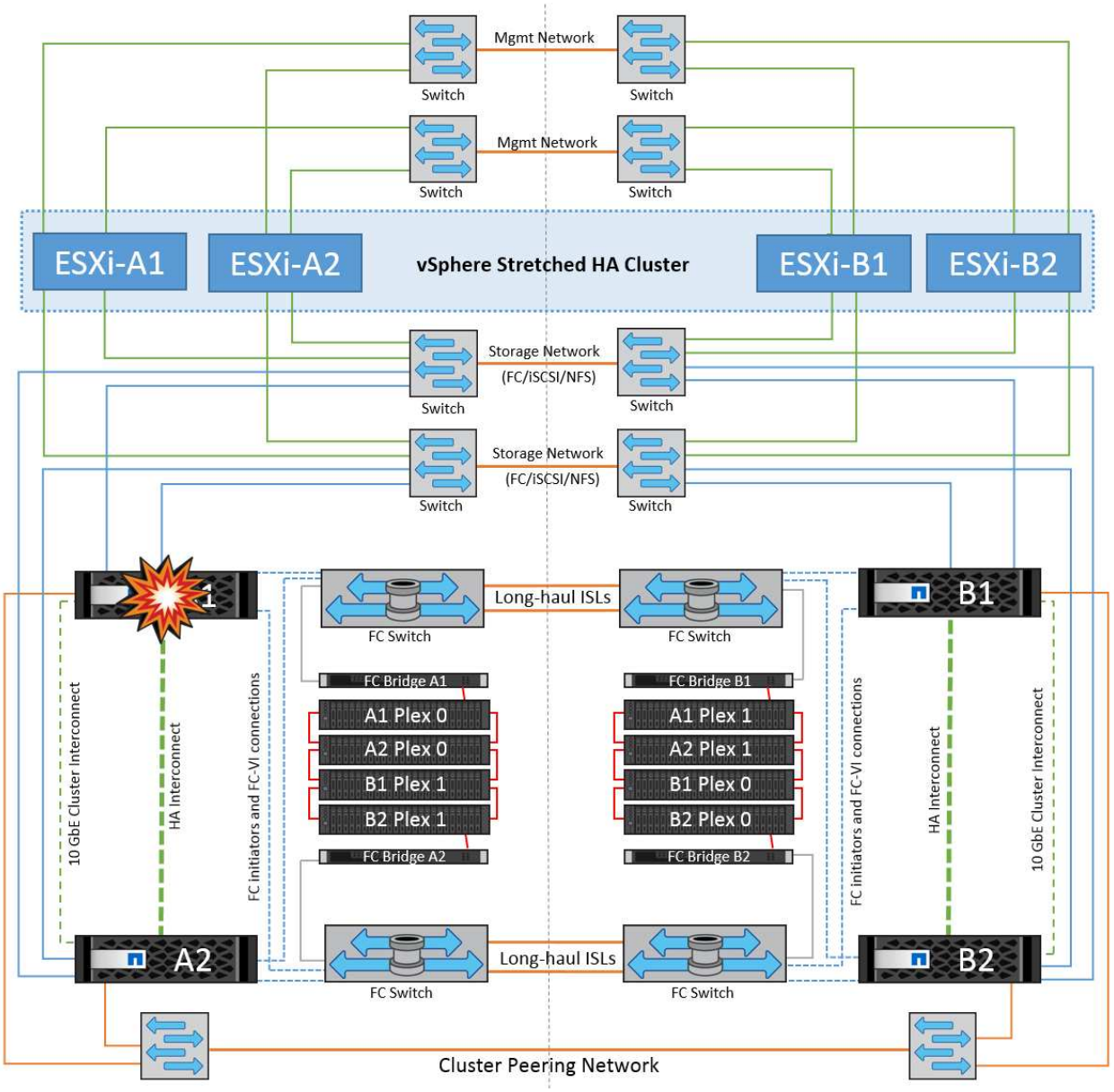
En esta situación, se produce un fallo de más de dos discos o una bandeja entera. Los datos se sirven desde el plex superviviente sin interrupción de los servicios de datos. El fallo del disco puede afectar a un plex local o remoto. Los agregados se mostrarán como degradado porque solo está activo un plex. Una vez sustituidos los discos que han fallado, los agregados afectados se sincronizarán automáticamente para volver a compilar los datos. Tras realizar la resincronización, los agregados volverán automáticamente al modo reflejado normal. Si ha fallado más de dos discos dentro de un mismo grupo RAID, es necesario reconstruir el plex desde cero.



Nota: Durante este período, no hay impacto en las operaciones de E/S de la máquina virtual, pero hay un rendimiento degradado porque se accede a los datos desde la bandeja de discos remotos a través de enlaces ISL.

Fallo de una controladora de almacenamiento única

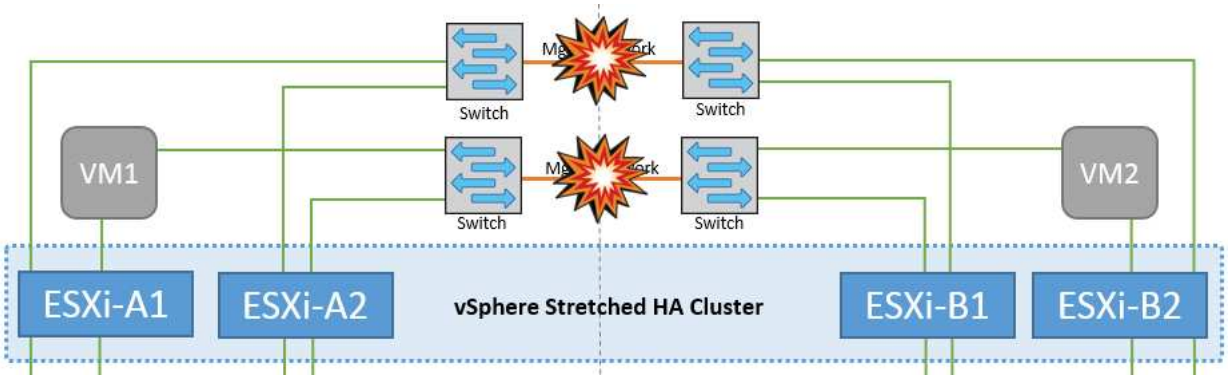
En este escenario, una de las dos controladoras de almacenamiento falla en un sitio. Dado que hay un par de alta disponibilidad en cada sitio, el fallo de un nodo de forma transparente activa automáticamente la conmutación al otro nodo. Por ejemplo, si falla el nodo A1, su almacenamiento y sus cargas de trabajo se transfieren automáticamente al nodo A2. Las máquinas virtuales no se verán afectadas porque todos los plexes permanecen disponibles. Los nodos del segundo sitio (B1 y B2) no se ven afectados. Además, vSphere HA no realizará ninguna acción porque el nodo principal del clúster seguirá recibiendo los latidos de red.



Si la conmutación al respaldo forma parte de un desastre gradual (el nodo A1 conmuta al nodo A2) y hay un fallo posterior de A2 o el fallo completo del sitio A, el cambio tras un desastre puede ocurrir en el sitio B.

Fallos de enlace de interinterruptor

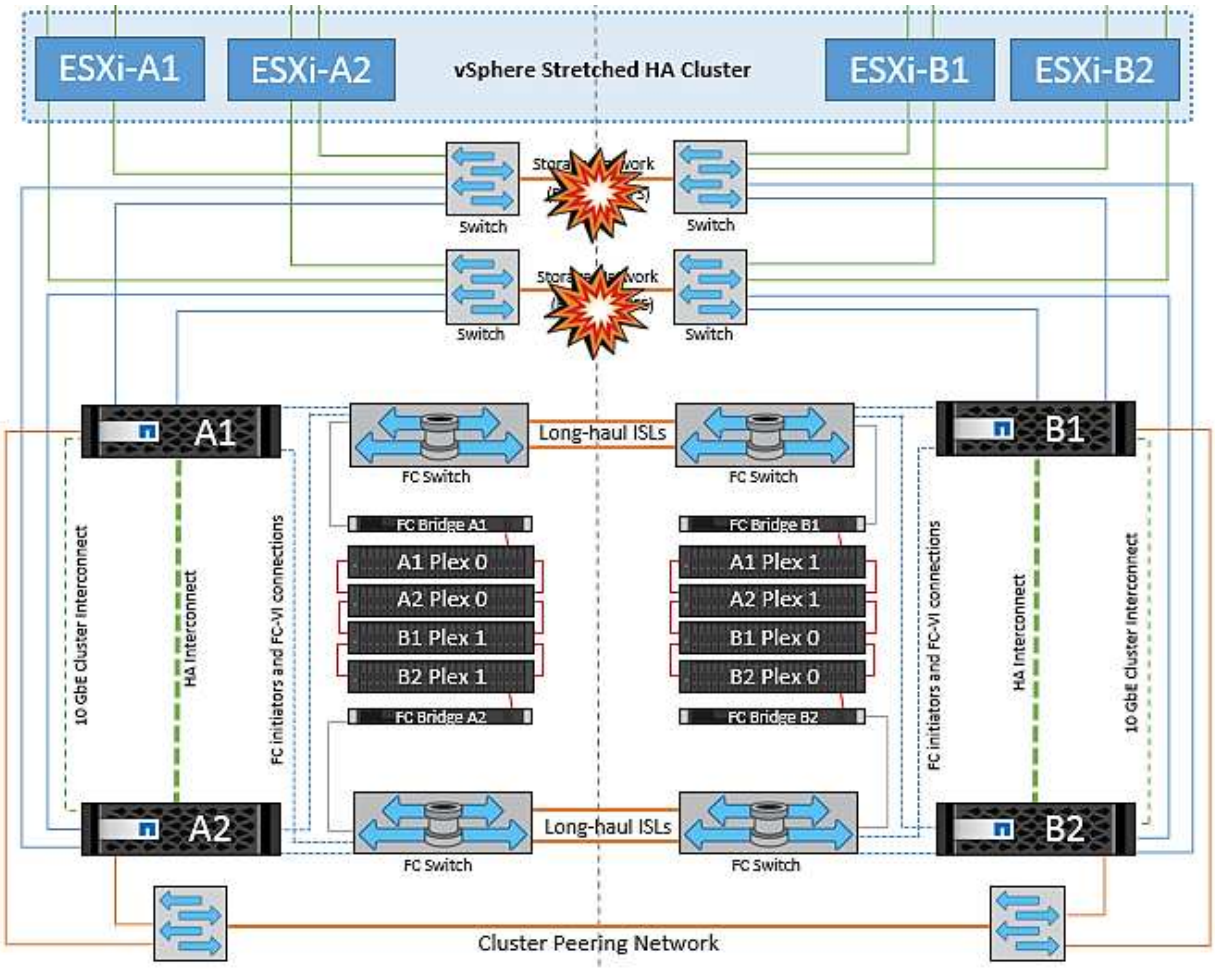
Fallo de enlace de interswitch en la red de gestión



En este escenario, si los enlaces ISL en la red de gestión de host de interfaz de usuario producen un error, los hosts ESXi del sitio A no podrán comunicarse con los hosts ESXi del sitio B. Esto dará lugar a una partición de red porque los hosts ESXi de un sitio concreto no podrán enviar los latidos de red al nodo maestro del clúster HA. Como tal, habrá dos segmentos de red debido a la partición y habrá un nodo maestro en cada segmento que protegerá las VM de fallos de host dentro del sitio en particular.

Nota: Durante este período, las máquinas virtuales permanecen en ejecución y no hay cambios en el comportamiento de MetroCluster en este escenario. Todos los almacenes de datos siguen estando intactos en sus respectivos sitios.

Fallo de enlace interswitch en la red de almacenamiento

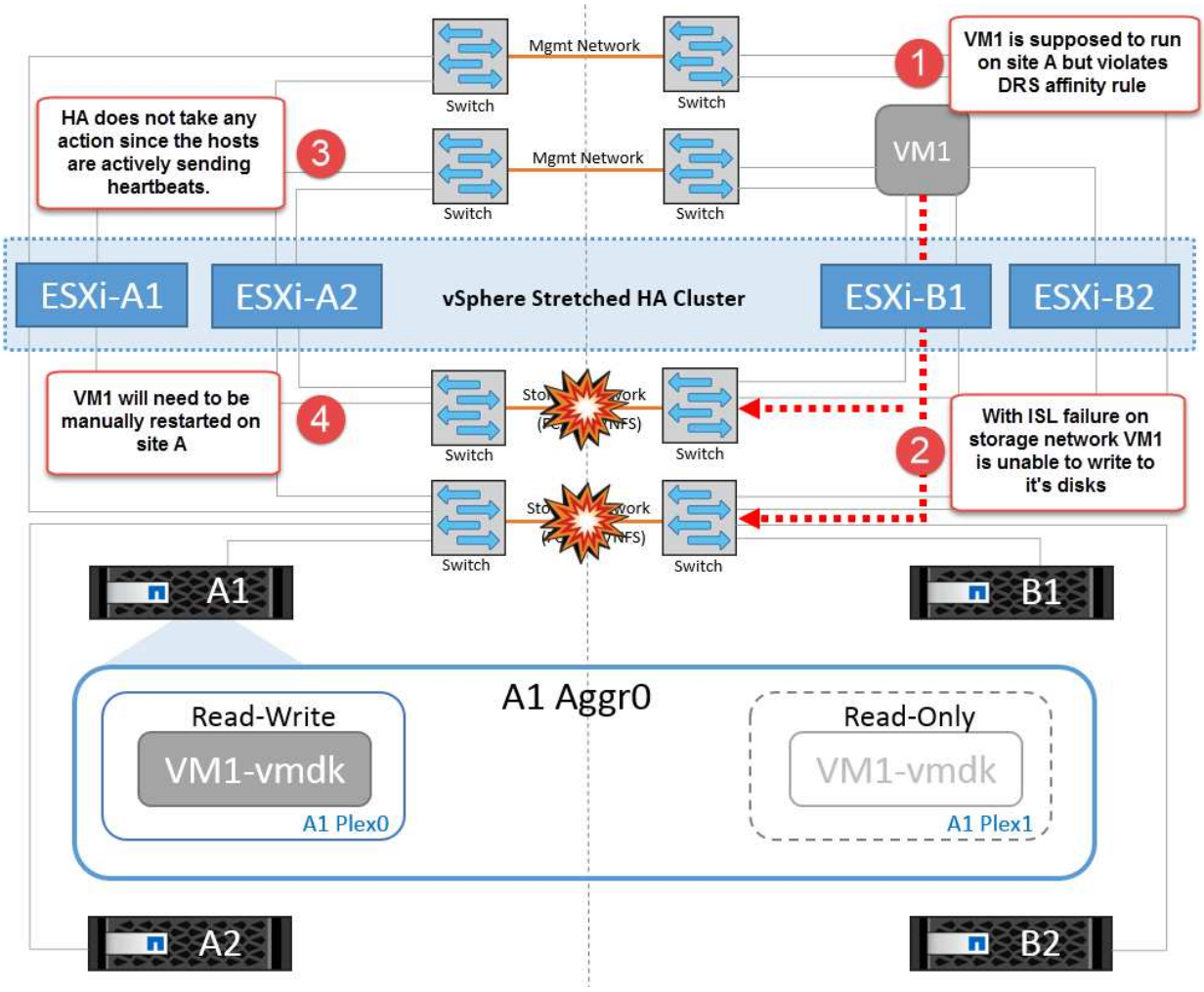


En este escenario, si los enlaces ISL en la red de almacenamiento de back-end fallan, los hosts del sitio A perderán acceso a los volúmenes de almacenamiento o las LUN del clúster B en el sitio B y viceversa. Las reglas de VMware DRS se definen de modo que la afinidad de sitios de almacenamiento host facilita que los equipos virtuales funcionen sin que el sitio se vea afectado.

Durante este período, las máquinas virtuales permanecen en ejecución en sus respectivos sitios y no hay cambios en el comportamiento de MetroCluster en este escenario. Todos los almacenes de datos siguen estando intactos en sus respectivos sitios.

Si por algún motivo se violó la regla de afinidad (por ejemplo, VM1, que se suponía que se ejecutaba desde la ubicación A donde sus discos residen en nodos del clúster local A, se está ejecutando en un host del sitio B), se accederá al disco de la máquina virtual de forma remota a través de enlaces ISL. Debido a un fallo de enlace ISL, VM1 ejecutándose en la instalación B no podría escribir en sus discos porque las rutas al volumen de almacenamiento están inactivas y la máquina virtual determinada está inactiva. En estos casos, VMware

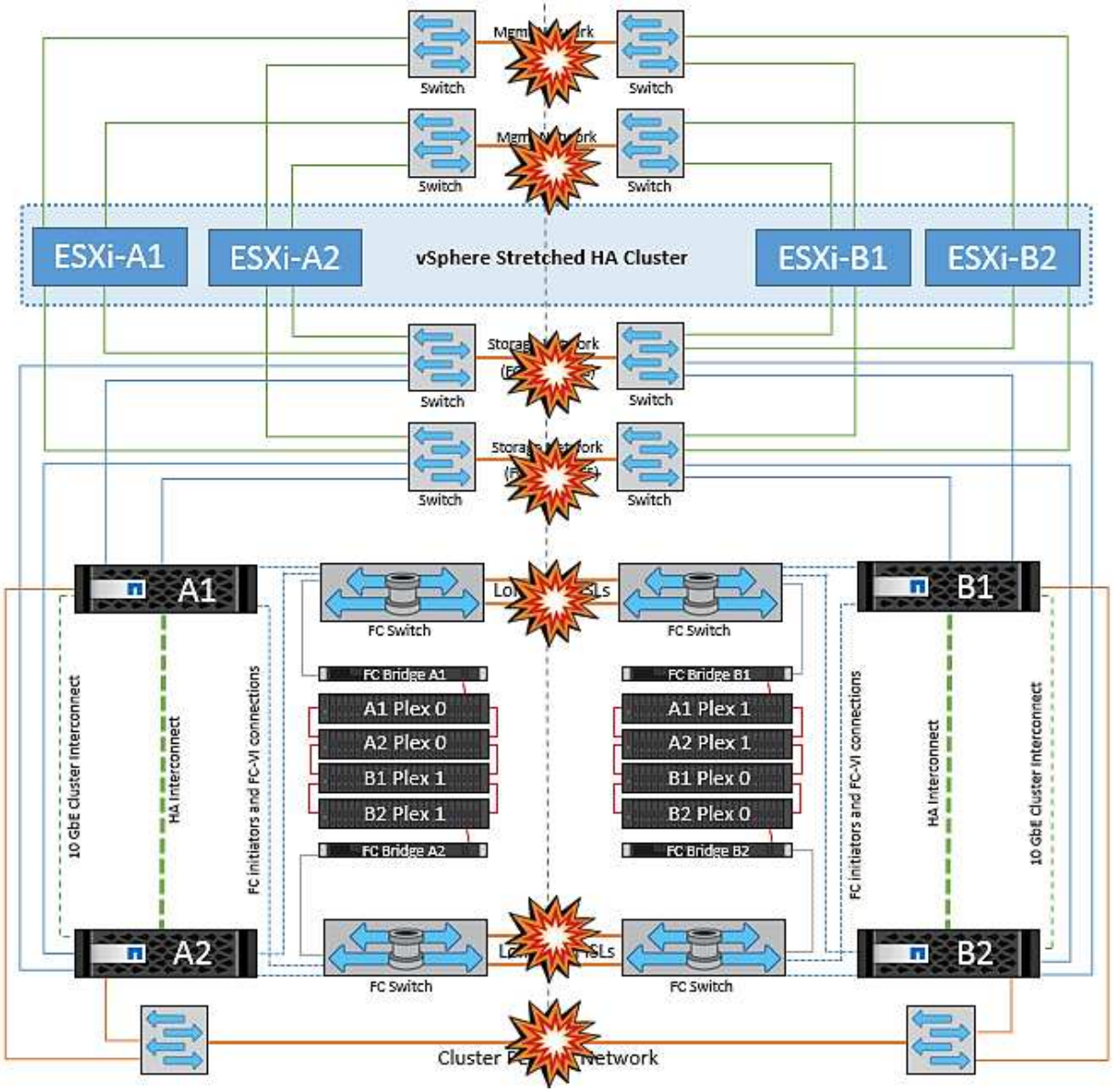
HA no realiza ninguna acción puesto que los hosts envían latidos de forma activa. Esas máquinas virtuales deben apagarse y encenderse manualmente en sus respectivos sitios. La siguiente figura ilustra una VM que viola una regla de afinidad DRS.



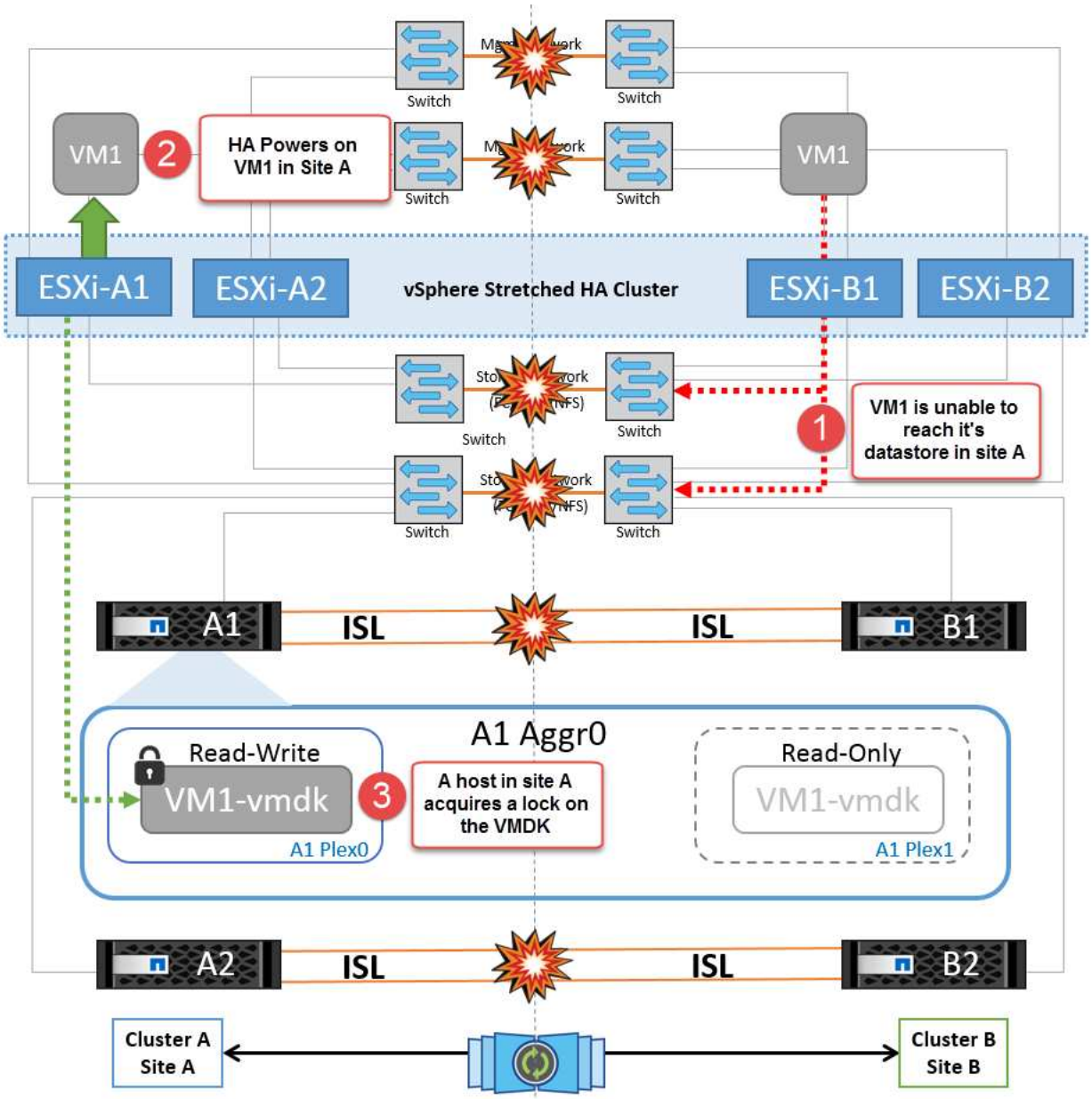
Todos los fallos de interswitch o la partición completa del centro de datos

En este escenario, todos los enlaces ISL entre los sitios están inactivos y los dos sitios están aislados uno de otro. Como se explicó en escenarios anteriores, como el fallo ISL en la red de gestión y en la red de almacenamiento, las máquinas virtuales no se ven afectadas por un fallo de ISL completo.

Una vez que los hosts ESXi hayan particionado entre sitios, el agente de alta disponibilidad de vSphere comprobará si hay latidos del almacén de datos y, en cada sitio, los hosts ESXi locales podrán actualizar los latidos del almacén de datos a sus respectivos volúmenes/LUN de lectura/escritura. Los hosts del sitio A asumirán que los otros hosts ESXi del sitio B han fallado porque no hay latidos de red/almacén de datos. La alta disponibilidad de vSphere en el sitio A intentará reiniciar las máquinas virtuales del sitio B, lo cual fallará en algún momento porque no se podrá acceder a los almacenes de datos del sitio B debido a un fallo del ISL de almacenamiento. Una situación similar se repite en el sitio B.



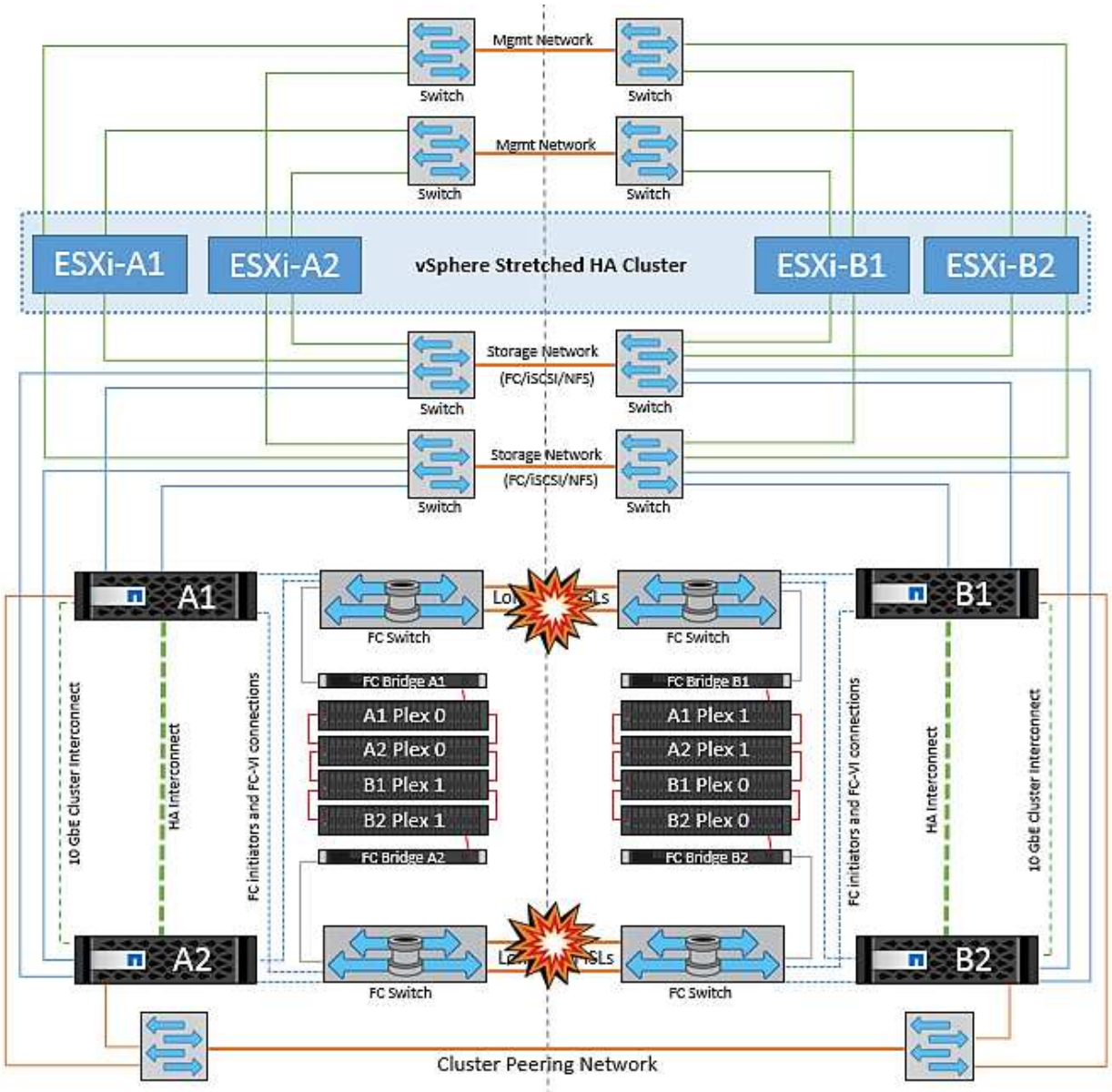
NetApp recomienda determinar si alguna máquina virtual ha infringido las reglas de DRS. Los equipos virtuales que se ejecuten desde un sitio remoto estarán inactivos ya que no podrán acceder al almacén de datos y vSphere HA reiniciará esa máquina virtual en el sitio local. Una vez que los enlaces ISL vuelvan a estar en línea, la máquina virtual que se estaba ejecutando en el sitio remoto se desactivará, ya que no puede haber dos instancias de máquinas virtuales ejecutándose con las mismas direcciones MAC.



Fallo de interswitch Link en ambas estructuras en NetApp MetroCluster

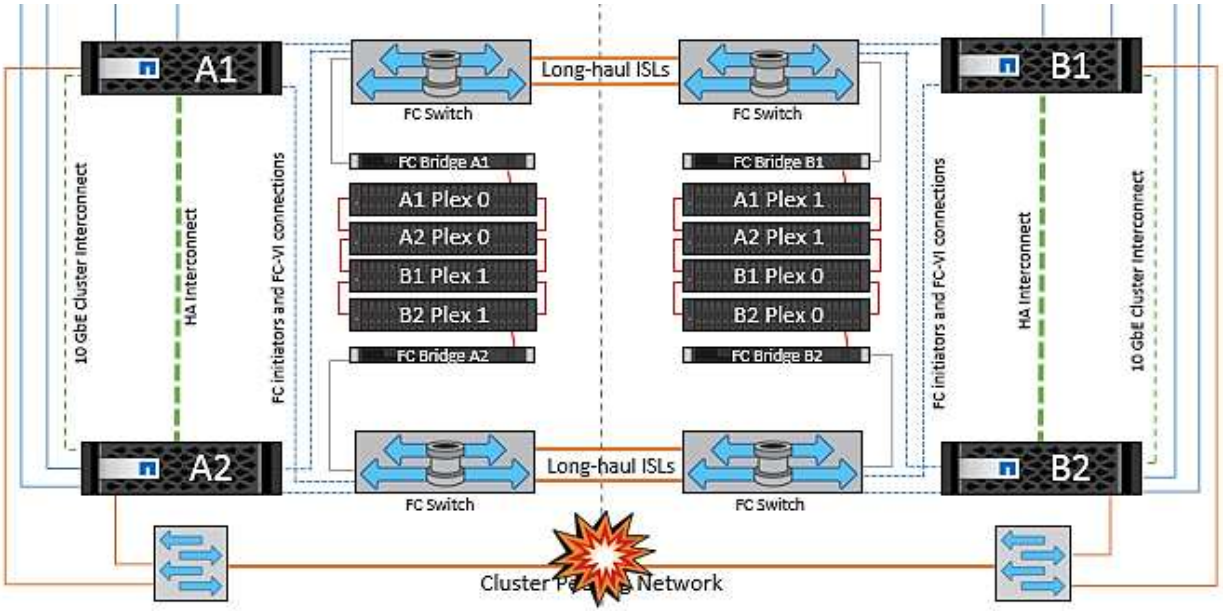
En un escenario en el que uno o varios ISL fallan, el tráfico continúa por los enlaces restantes. Si todos los ISL de ambas estructuras fallan, de modo que no hay ningún enlace entre los sitios para el almacenamiento y la replicación de NVRAM, cada controladora seguirá proporcionando sus datos locales. Al restaurar un mínimo de un ISL, la resincronización de todos los complejos se realizará automáticamente.

Las escrituras que se produzcan después de que todos los ISL estén inactivos no se reflejarán en el otro sitio. Una conmutación de sitios en caso de desastre, mientras la configuración se encuentra en este estado, por lo tanto, incurriría en la pérdida de los datos que no se habían sincronizado. En este caso, se requiere intervención manual para la recuperación después del cambio. Si es probable que no haya ISL disponibles durante un largo período de tiempo, un administrador puede optar por cerrar todos los servicios de datos para evitar el riesgo de pérdida de datos si es necesario una conmutación por desastre. La realización de esta acción debe evaluarse para la probabilidad de que se produzca un desastre que requiera la conmutación del servicio antes de que esté disponible al menos un ISL. Como alternativa, si los ISL fallan en un escenario en cascada, un administrador podría activar una conmutación de sitios planificada a uno de los sitios antes de que todos los enlaces hayan fallado.



Fallo de enlace de clúster con conexión entre iguales

En un supuesto de fallo de enlace de clústeres con conexión entre iguales, dado que los ISL de estructura aún están activos, los servicios de datos (lecturas y escrituras) continúan en ambos sitios en ambos complejos. No se puede propagar ningún cambio de configuración del clúster (por ejemplo, añadir una nueva SVM o aprovisionar un volumen o un LUN en una SVM existente) al otro sitio. Estos se mantienen en los volúmenes de metadatos de CRS locales y se propagan automáticamente al otro clúster al restaurar el enlace de clúster entre iguales. Si se necesita una conmutación por error forzada antes de poder restaurar el enlace de clúster entre iguales, se volverán a reproducir automáticamente los cambios pendientes de configuración de clúster desde la copia replicada remota de los volúmenes de metadatos del sitio superviviente como parte del proceso de conmutación por error.



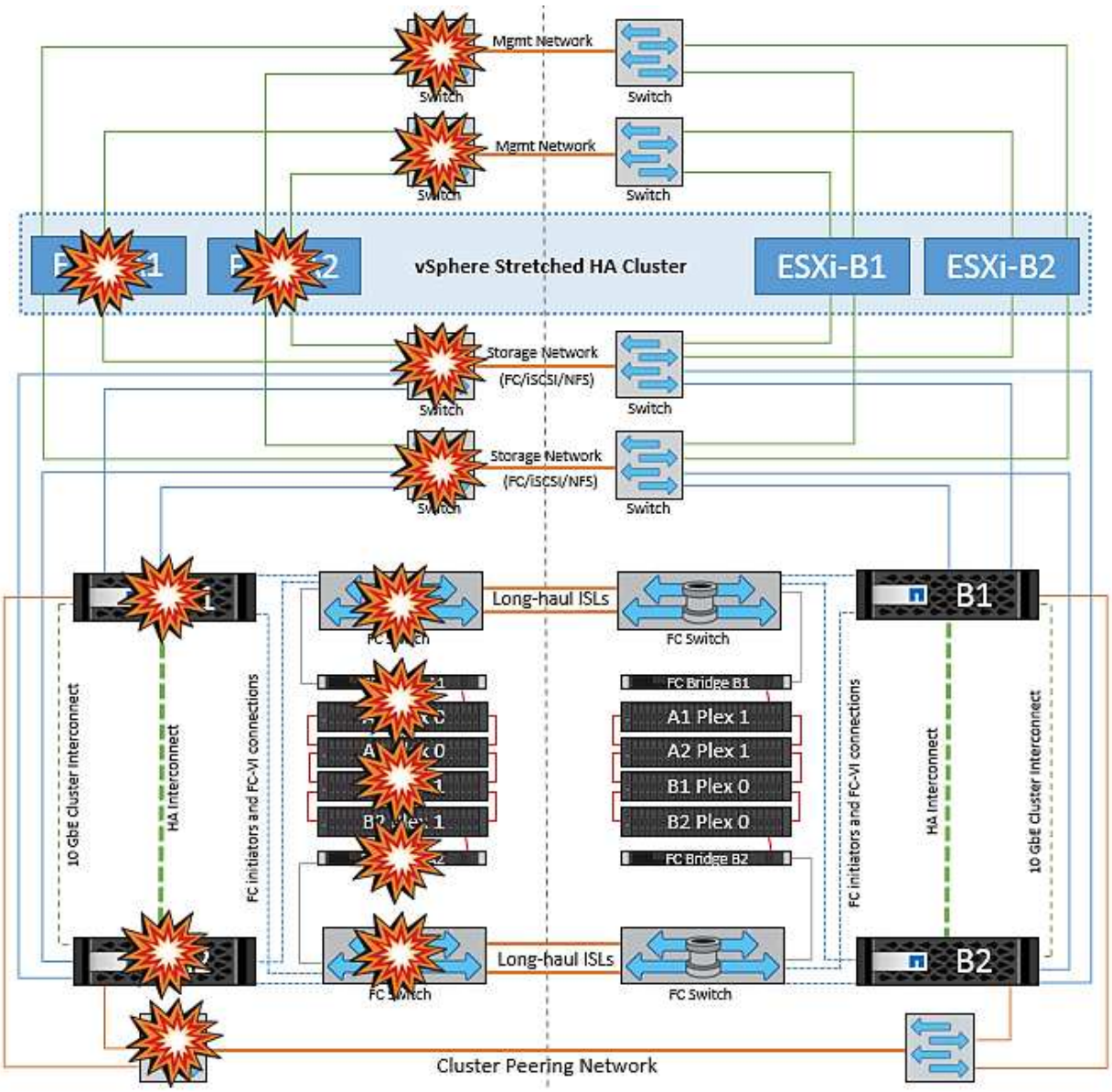
Fallo completo del sitio

En un supuesto de fallo del sitio A completo, los hosts ESXi del sitio B no obtendrán el latido de red de los hosts ESXi del sitio A porque están inactivos. El maestro de alta disponibilidad en el sitio B verificará que los latidos del almacén de datos no están presentes, declarará que los hosts del sitio A han fallado e intentará reiniciar el sitio A de los equipos virtuales en el sitio B. Durante este periodo, el administrador de almacenamiento realiza una conmutación de sitios para reanudar los servicios de los nodos fallidos en el sitio superviviente. Esto restaura todos los servicios de almacenamiento del sitio A en el sitio B. Después de que el sitio haya volúmenes o LUN disponibles en el sitio B, el agente maestro de alta disponibilidad intentará reiniciar el sitio A, máquinas virtuales del sitio B.

Si el intento del agente maestro HA de vSphere de reiniciar una máquina virtual (lo que implica registrarla y encenderla) falla, el reinicio se vuelve a intentar después de un retraso. El retardo entre reinicios se puede configurar hasta un máximo de 30 minutos. VSphere HA intenta estos reinicios durante un número máximo de intentos (seis intentos de forma predeterminada).

Nota: El maestro HA no inicia los intentos de reinicio hasta que el administrador de colocación encuentre el almacenamiento adecuado, por lo que en el caso de un fallo completo del sitio, eso sería después de que se haya realizado el cambio.

Si el sitio A se ha cambiado, un fallo posterior de uno de los nodos del sitio B superviviente se puede gestionar sin problemas mediante la conmutación al nodo superviviente. En este caso, solo un nodo realiza el trabajo de cuatro nodos. En este caso, la recuperación consistiría en realizar un retorno al nodo local. A continuación, cuando se restaura el sitio A, se realiza una operación de conmutación para restaurar el funcionamiento en estado constante de la configuración.


Mgmt Network
Switch
Mgmt Network
Switch
vSphere Stretched HA Cluster
ESXi-B1
ESXi-B2
Storage Network
(FC/iSCSI/NFS)
Switch
Storage Network
(FC/iSCSI/NFS)
Switch
Long-haul ISLs
FC Switch
B1
FC Bridge B1
A1 Plex 1
A2 Plex 1
B1 Plex 0
B2 Plex 0
FC Bridge B2
10 GbE Cluster Interconnect
HA Interconnect
FC initiators and FC-VI connections
FC initiators and FC-VI connections
HA Interconnect
10 GbE Cluster Interconnect
Long-haul ISLs
FC Switch
B2
Cluster Peering Network