



SnapMirror Active Sync

Enterprise applications

NetApp

February 10, 2026

This PDF was generated from <https://docs.netapp.com/de-de/ontap-apps-dbs/oracle/oracle-dr-smas-overview.html> on February 10, 2026. Always check docs.netapp.com for the latest.

Inhalt

SnapMirror Active Sync	1
Überblick	1
Synchrone Replizierung	1
Storage-Hardware	1
ONTAP Mediator	1
ONTAP Mediator	1
Bevorzugter Standort für SnapMirror Active Sync	3
Netzwerktopologie	4
Einheitlicher Zugriff	4
Uneinheitlicher Zugriff	9
Oracle Konfigurationen	11
Überblick	11
Oracle Single Instance	11
Oracle Extended RAC	13
RAC Tiebreaker	22
Ausfallszenarien	23
Überblick	23
Beispielarchitektur	25
RAC-Verbindungsfehler	26
SnapMirror-Kommunikationsfehler	27
Totaler Fehler bei der Netzwerkverbindung	28
Standortausfall	30
Fehler beim Mediator	32
Servicewiederherstellung	33
Manueller Failover	33

SnapMirror Active Sync

Überblick

Mit SnapMirror Active Sync können Sie Oracle-Datenbankumgebungen mit extrem hoher Verfügbarkeit aufbauen, in denen LUNs von zwei verschiedenen Storage-Clustern verfügbar sind.

Bei SnapMirror Active Sync gibt es keine „primäre“ und „sekundäre“ Kopie der Daten. Jedes Cluster kann Lese-I/O aus seiner lokalen Kopie der Daten bereitstellen, und jedes Cluster repliziert einen Schreibvorgang auf seinen Partner. Das Ergebnis ist ein symmetrisches I/O-Verhalten.

So können Sie unter anderem Oracle RAC als erweiterten Cluster mit operativen Instanzen an beiden Standorten ausführen. Alternativ können Sie RPO=0 aktiv/Passiv-Datenbank-Cluster erstellen, bei denen Datenbanken mit einer Instanz bei einem Standortausfall zwischen Standorten verschoben werden können. Dieser Prozess kann über Produkte wie Pacemaker oder VMware HA automatisiert werden. Die Grundlage für all diese Optionen ist die synchrone Replizierung, die über SnapMirror Active Sync gemanagt wird.

Synchrone Replizierung

Im normalen Betrieb bietet SnapMirror Active Sync jederzeit ein synchrones RPO=0-Replikat, mit einer Ausnahme. Wenn Daten nicht repliziert werden können, gibt ONTAP die Notwendigkeit zur Replizierung von Daten frei und stellt die E/A-Bereitstellung an einem Standort wieder her, während die LUNs am anderen Standort offline geschaltet werden.

Storage-Hardware

Im Gegensatz zu anderen Disaster Recovery-Lösungen für Storage bietet SnapMirror Active Sync asymmetrische Plattformflexibilität. Die Hardware an den einzelnen Standorten muss nicht identisch sein. Dank dieser Funktion können Sie die Größe der Hardware anpassen, die zur Unterstützung der SnapMirror Active Sync verwendet wird. Das Remote-Storage-System kann identisch mit dem primären Standort sein, wenn es einen vollständigen Produktions-Workload unterstützen muss. Wenn jedoch ein Ausfall zu einer Verringerung der I/O führt, könnte ein kleineres System am Remote-Standort kostengünstiger sein.

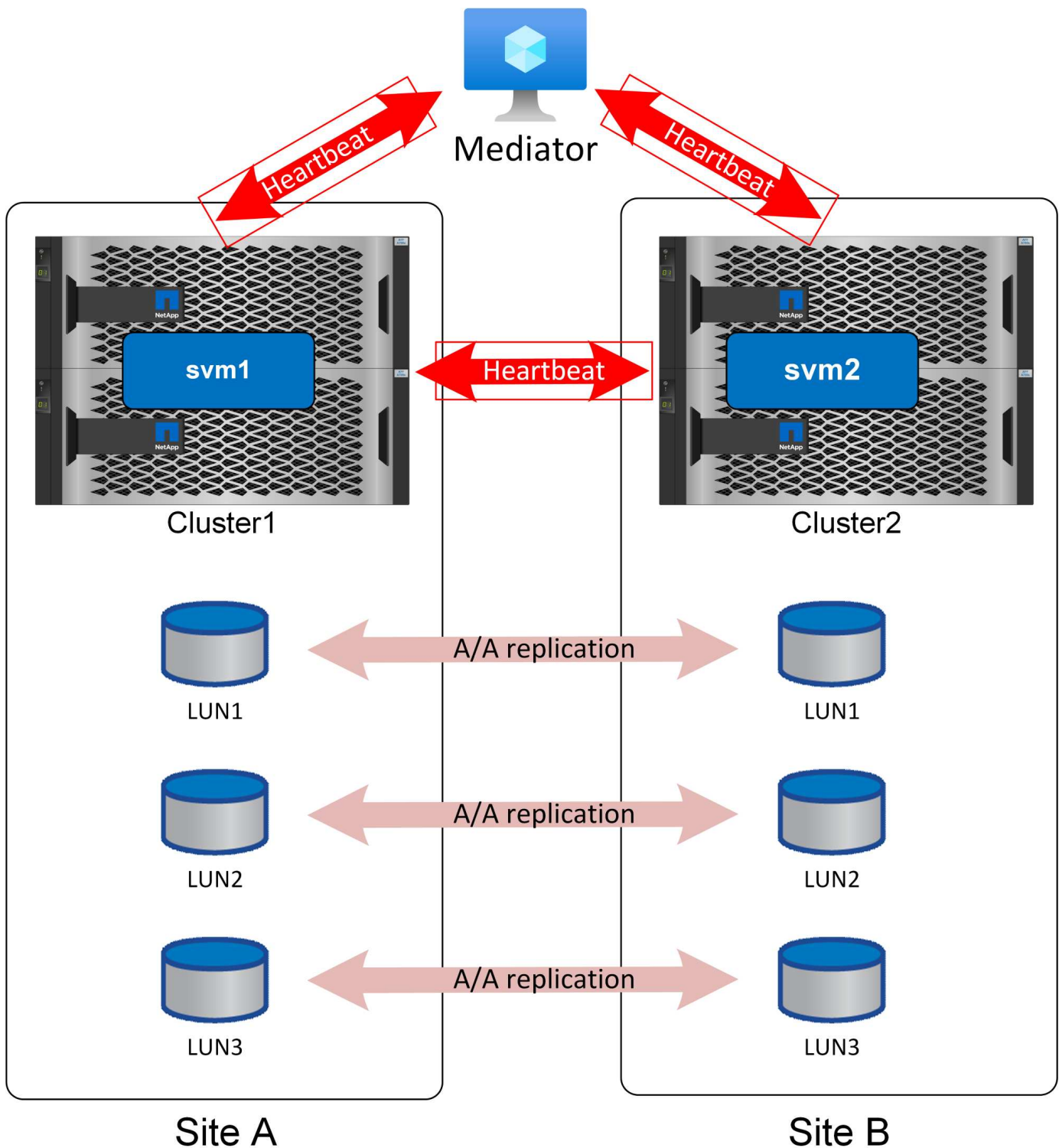
ONTAP Mediator

Der ONTAP Mediator ist eine Softwareanwendung, die von der NetApp-Unterstützung heruntergeladen wird und normalerweise auf einer kleinen virtuellen Maschine bereitgestellt wird. Bei Verwendung mit SnapMirror Active Sync ist der ONTAP Mediator kein Tiebreaker. Es handelt sich um einen alternativen Kommunikationskanal für die beiden Cluster, die an der aktiven synchronen SnapMirror-Replikation beteiligt sind. Der automatisierte Betrieb wird durch ONTAP basierend auf den Antworten gesteuert, die der Partner über direkte Verbindungen und den Mediator erhält.

ONTAP Mediator

Der Mediator ist für die sichere Automatisierung des Failover erforderlich. Idealerweise würde er an einem unabhängigen dritten Standort platziert werden, kann aber dennoch für die meisten Anforderungen funktionieren, wenn er mit einem der an der Replikation beteiligten Cluster kolokiert wird.

Der Mediator ist eigentlich kein Entscheider bei Stimmengleichständen, obwohl er diese Funktion faktisch übernimmt. Der Mediator hilft dabei, den Zustand der Clusterknoten zu ermitteln und unterstützt den automatischen Umschaltprozess im Falle eines Standortausfalls. Der Mediator übermittelt unter keinen Umständen Daten.



Die #1 Herausforderung mit automatisiertem Failover ist das Split-Brain-Problem, und dieses Problem tritt auf, wenn Ihre zwei Standorte die Verbindung miteinander verlieren. Was soll geschehen? Sie möchten nicht, dass sich zwei verschiedene Standorte als verbleibende Kopien der Daten bezeichnen, aber wie kann ein einzelner Standort den Unterschied zwischen dem tatsächlichen Verlust des anderen Standorts und der Unfähigkeit, mit

dem gegenüberliegenden Standort zu kommunizieren, erkennen?

Hier betritt der Mediator das Bild. Wenn jeder Standort an einem dritten Standort platziert wird und über eine separate Netzwerkverbindung zu diesem Standort verfügt, haben Sie für jeden Standort einen zusätzlichen Pfad, um den Zustand des anderen zu überprüfen. Sehen Sie sich das Bild oben noch einmal an und betrachten Sie die folgenden Szenarien.

- Was passiert, wenn der Mediator ausfällt oder von einem oder beiden Standorten nicht erreichbar ist?
 - Die beiden Cluster können weiterhin über dieselbe Verbindung miteinander kommunizieren, die für Replikationsdienste verwendet wird.
 - Für die Daten wird noch eine RPO=0-Sicherung verwendet
- Was passiert, wenn Standort A ausfällt?
 - An Standort B sehen Sie, dass beide Kommunikationskanäle ausgefallen sind.
 - Standort B übernimmt die Datenservices, jedoch ohne RPO=0-Spiegelung
- Was passiert, wenn Standort B ausfällt?
 - An Standort A sehen Sie, dass beide Kommunikationskanäle ausgefallen sind.
 - Standort A übernimmt die Datenservices, aber ohne RPO=0-Spiegelung

Es gibt ein anderes Szenario zu berücksichtigen: Verlust der Datenreplikationsverbindung. Wenn die Replikationsverbindung zwischen Standorten verloren geht, wird eine RPO=0-Spiegelung offensichtlich unmöglich sein. Was soll dann geschehen?

Dies wird durch den bevorzugten Standortstatus gesteuert. In einer SM-AS-Beziehung ist einer der Standorte zweitrangig zum anderen. Dies hat keine Auswirkungen auf den normalen Betrieb, und der gesamte Datenzugriff ist symmetrisch. Wenn die Replikation jedoch unterbrochen wird, muss die Verbindung unterbrochen werden, um den Betrieb wieder aufzunehmen. Das Ergebnis: Der bevorzugte Standort setzt den Betrieb ohne Spiegelung fort und der sekundäre Standort hält die I/O-Verarbeitung an, bis die Replizierungskommunikation wiederhergestellt ist.

Bevorzugter Standort für SnapMirror Active Sync

Das aktive Synchronisierungsverhalten von SnapMirror ist symmetrisch, mit einer wichtigen Ausnahme: Konfiguration des bevorzugten Standorts.

SnapMirror Active Sync betrachtet einen Standort als „Quelle“ und den anderen als „Ziel“. Dies impliziert eine One-Way-Replikationsbeziehung, aber dies gilt nicht für das IO-Verhalten. Die Replizierung ist bidirektional und symmetrisch, und die I/O-Reaktionszeiten sind auf beiden Seiten der Spiegelung identisch.

Die `source` Bezeichnung steuert den bevorzugten Standort. Wenn die Replizierungsverbindung verloren geht, stellen die LUN-Pfade auf der Quellkopie weiterhin Daten bereit, während die LUN-Pfade auf der Zielkopie erst dann wieder verfügbar sind, wenn die Replikation wiederhergestellt ist und SnapMirror wieder in den synchronen Zustand wechselt. Die Pfade setzen dann das Bereitstellen von Daten fort.

Die Sourcing/Ziel-Konfiguration kann über Systemmanager angezeigt werden:

Relationships		
Local destinations		Local sources
<div> <div>Search</div> <div>Download</div> <div>Show/hide: ▼</div> <div>Filter</div> </div>		
Source	Destination	Policy type
<div>▼</div> jfs_as1:/cg/jfsAA	jfs_as2:/cg/jfsAA	Synchronous

Oder über die CLI:

```
Cluster2::> snapmirror show -destination-path jfs_as2:/cg/jfsAA

Source Path: jfs_as1:/cg/jfsAA
Destination Path: jfs_as2:/cg/jfsAA
Relationship Type: XDP
Relationship Group Type: consistencygroup
SnapMirror Schedule: -
SnapMirror Policy Type: automated-failover-duplex
SnapMirror Policy: AutomatedFailOverDuplex
Tries Limit: -
Throttle (KB/sec): -
Mirror State: Snapmirrored
Relationship Status: InSync
```

Der Schlüssel ist, dass die Quelle die SVM für Cluster1 ist. Wie oben erwähnt, beschreiben die Begriffe „Quelle“ und „Ziel“ nicht den Fluss replizierter Daten. Beide Standorte können einen Schreibvorgang verarbeiten und am anderen Standort replizieren. Beide Cluster sind Quellen und Ziele. Der Effekt der Festlegung eines Clusters als Quelle steuert einfach, welches Cluster als Lese-/Schreib-Speichersystem überlebt, wenn die Replikationsverbindung verloren geht.

Netzwerktopologie

Einheitlicher Zugriff

Ein einheitliches Netzwerk für den Zugriff bedeutet, dass Hosts auf Pfade auf beiden Seiten (oder auf Ausfall-Domains innerhalb desselben Standorts) zugreifen können.

Eine wichtige Funktion von SM-AS ist die Möglichkeit, die Speichersysteme so zu konfigurieren, dass sie wissen, wo sich die Hosts befinden. Wenn Sie die LUNs einem bestimmten Host zuordnen, können Sie angeben, ob sie einem bestimmten Storage-System proximal sind oder nicht.

Annäherungseinstellungen

Proximity bezieht sich auf eine Clusterkonfiguration, die angibt, dass eine bestimmte Host-WWN- oder iSCSI-

Initiator-ID zu einem lokalen Host gehört. Dies ist ein zweiter optionaler Schritt für die Konfiguration des LUN-Zugriffs.

Der erste Schritt ist die übliche igroup-Konfiguration. Jede LUN muss einer Initiatorgruppe zugeordnet werden, die die WWN/iSCSI-IDs der Hosts enthält, die Zugriff auf diese LUN benötigen. Dadurch wird gesteuert, welcher Host *Access* zu einer LUN hat.

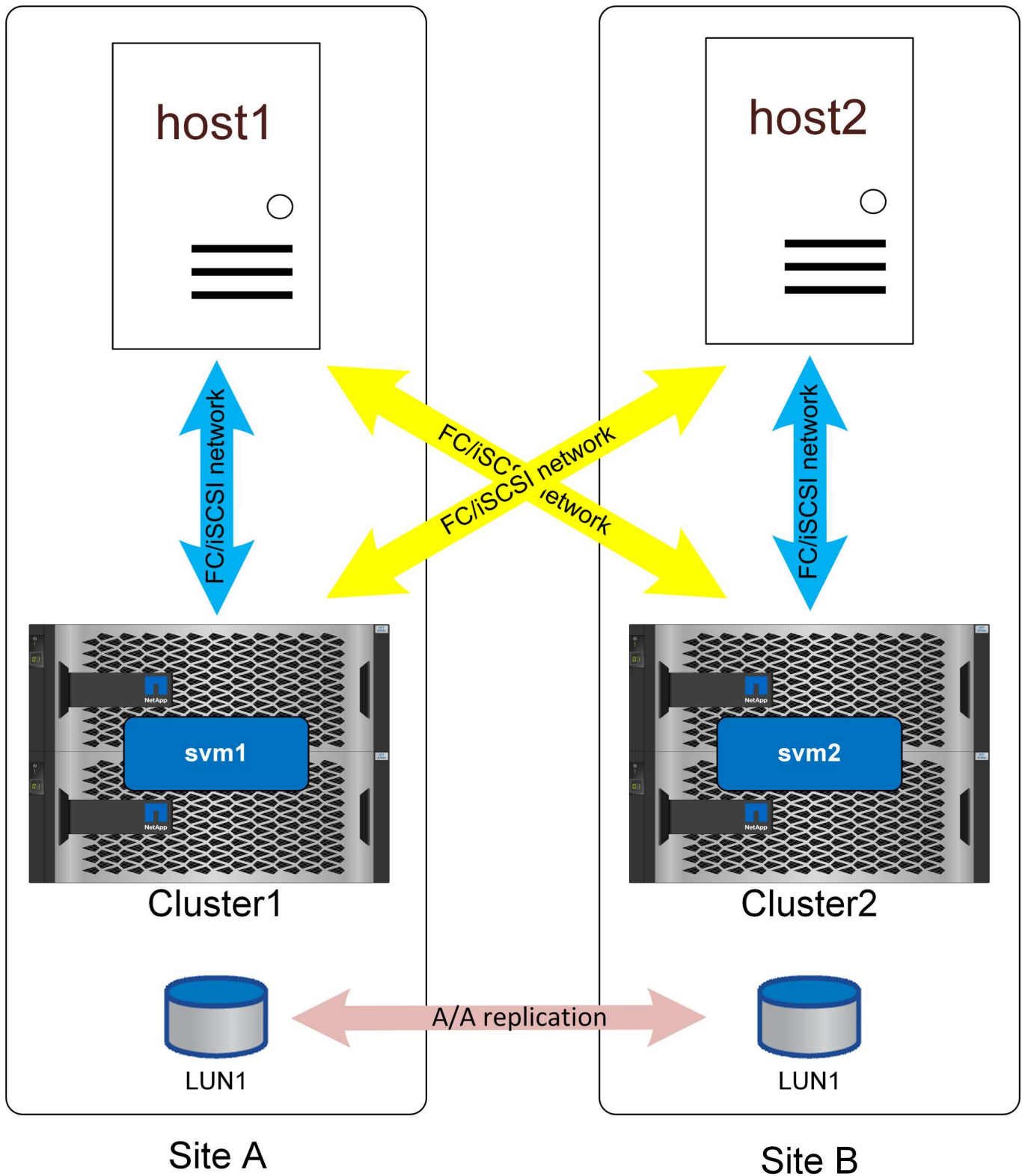
Der zweite, optionale Schritt ist die Konfiguration der Host-Nähe. Dies kontrolliert nicht den Zugriff, es steuert *Priority*.

Beispielsweise kann ein Host an Standort A für den Zugriff auf eine LUN konfiguriert werden, die durch SnapMirror Active Sync geschützt ist. Da das SAN über Standorte erweitert wird, stehen diesem LUN Pfade über Storage an Standort A oder Storage an Standort B zur Verfügung

Ohne Annäherungseinstellungen verwendet der Host beide Speichersysteme gleichmäßig, da beide Speichersysteme aktive/optimierte Pfade anbieten. Wenn die SAN-Latenz und/oder Bandbreite zwischen Standorten begrenzt ist, ist dies möglicherweise nicht erwünscht, und Sie sollten sicherstellen, dass während des normalen Betriebs jeder Host bevorzugt Pfade zum lokalen Speichersystem verwendet. Diese Konfiguration erfolgt durch Hinzufügen der Host-WWN/iSCSI-ID zum lokalen Cluster als proximaler Host. Dies kann unter der CLI oder Systemmanager ausgeführt werden.

AFF

Bei einem AFF System werden die Pfade nach dem Konfigurieren von Host-Nähe wie unten dargestellt angezeigt.



Active/Optimized Path

Active Path

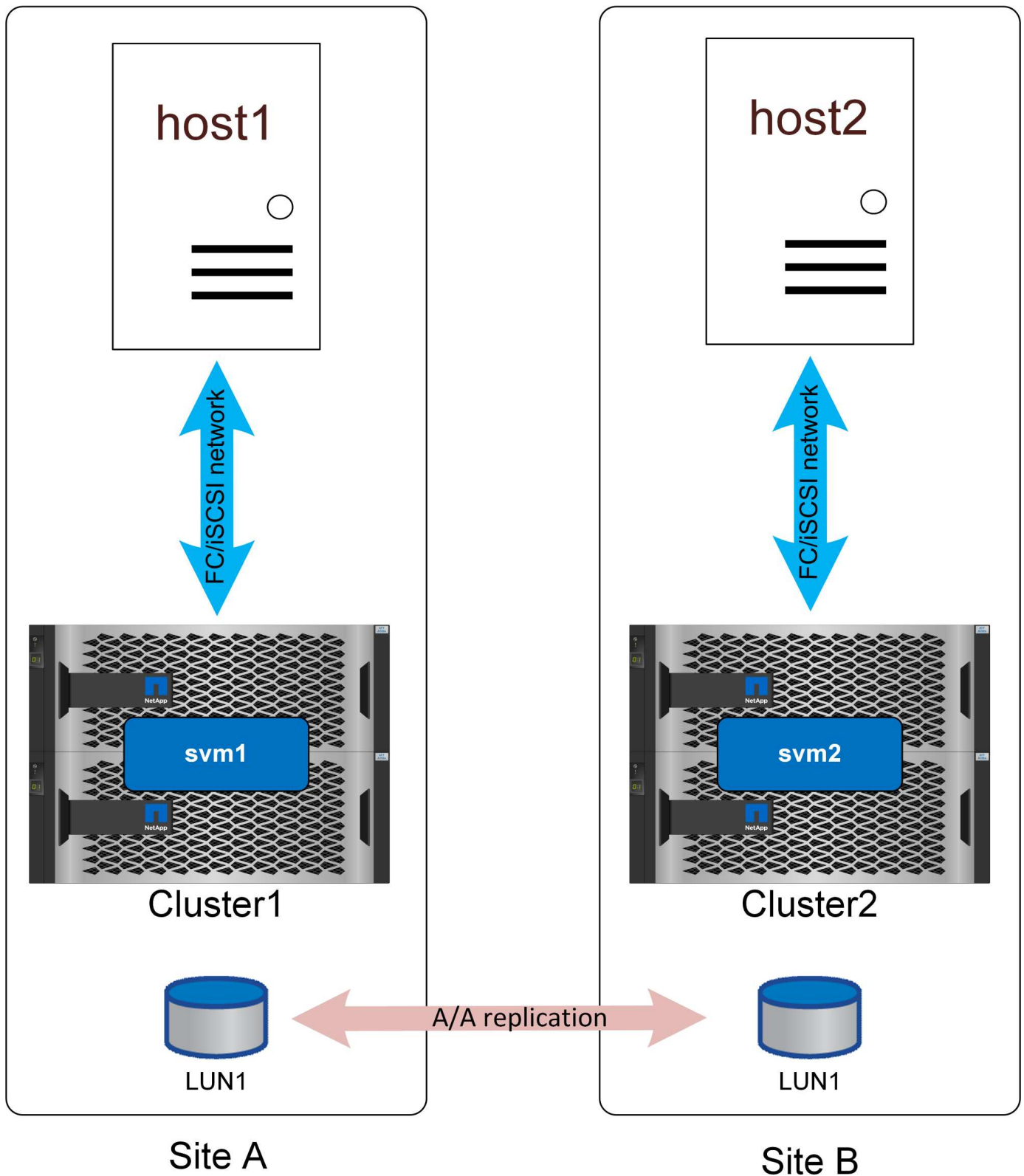
Im normalen Betrieb sind alle E/A-Vorgänge lokal. Lese- und Schreibvorgänge werden vom lokalen Speicher-Array gewartet. Schreib-I/O muss natürlich auch vom lokalen Controller auf das Remote-System repliziert werden, bevor sie bestätigt wird. Alle Lese-I/O-Vorgänge werden jedoch lokal gewartet und es kommt keine zusätzliche Latenz durch das Durchlaufen der SAN-Verbindung zwischen den Standorten zu.

Die nicht optimierten Pfade werden nur dann verwendet, wenn alle aktiven/optimierten Pfade verloren gehen. Wenn beispielsweise das gesamte Array an Standort A Strom verloren hätte, könnten die Hosts an Standort A weiterhin auf Pfade zum Array an Standort B zugreifen und bleiben daher betriebsbereit, obwohl die Latenz höher wäre.

Es gibt redundante Pfade durch den lokalen Cluster, die aus Gründen der Einfachheit nicht auf diesen Diagrammen angezeigt werden. ONTAP Storage-Systeme sind HA selbst, daher sollte ein Controller-Ausfall nicht zu einem Standortausfall führen. Es sollte lediglich zu einer Änderung führen, in der lokale Pfade auf dem betroffenen Standort verwendet werden.

ASA

NetApp ASA Systeme bieten aktiv/aktiv-Multipathing über alle Pfade eines Clusters hinweg. Dies gilt auch für SM-AS Konfigurationen.



Active/Optimized Path

Eine ASA-Konfiguration mit nicht-einheitlichem Zugriff würde im Wesentlichen auf die gleiche Weise funktionieren wie mit AFF. Bei einheitlichem Zugriff würde IO das WAN überqueren. Dies kann wünschenswert

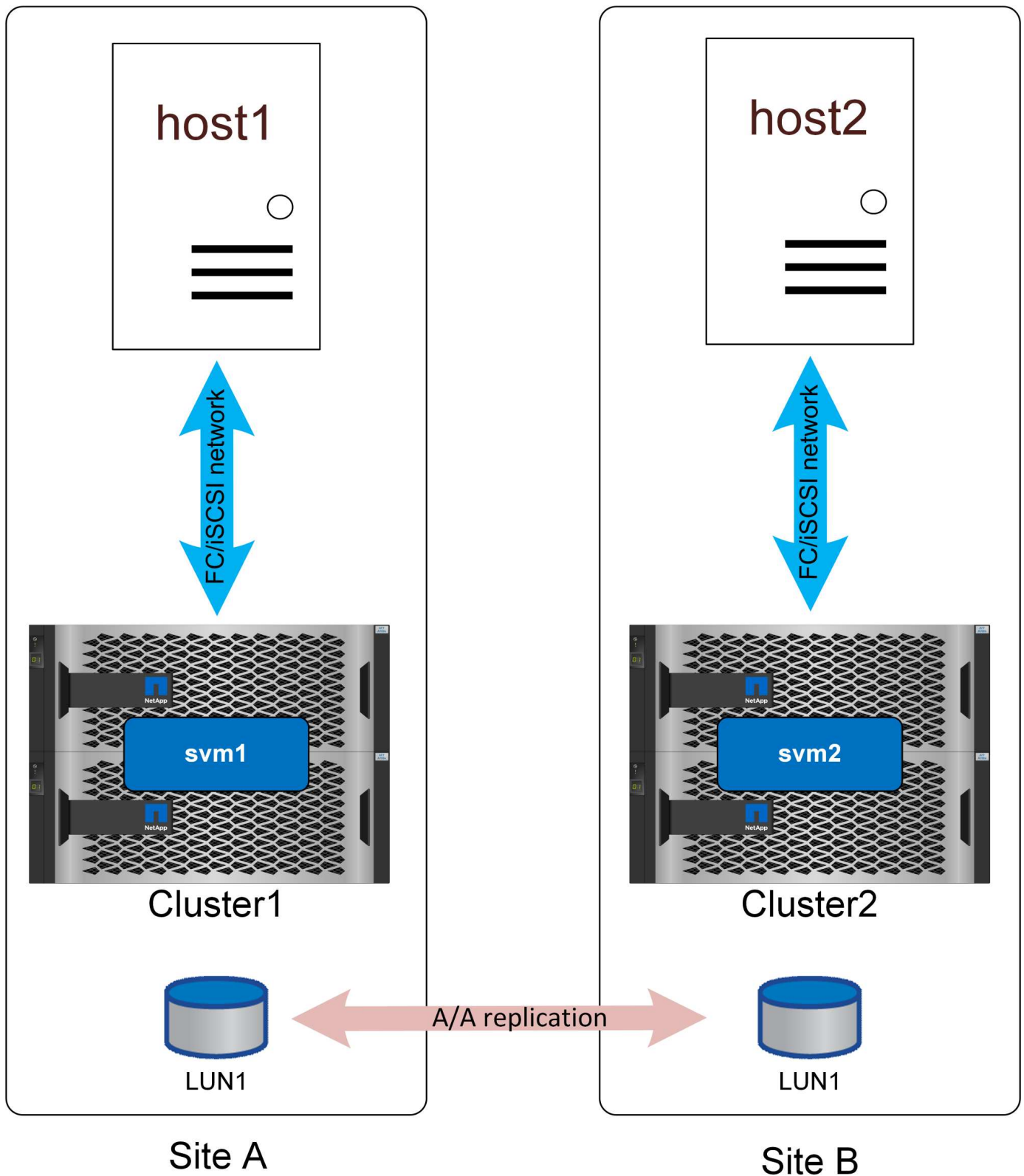
sein oder auch nicht.

Wenn die beiden Standorte mit Glasfaserverbindung 100 Meter voneinander entfernt wären, sollte keine erkennbare zusätzliche Latenz über das WAN entstehen. Wenn jedoch die Standorte weit voneinander entfernt wären, würde die Performance beim Lesen an beiden Standorten darunter leiden. Im Gegensatz dazu würden bei AFF diese WAN-überschneidenden Pfade nur verwendet, wenn keine lokalen Pfade verfügbar wären und die tägliche Performance besser wäre, da alle I/O-Vorgänge lokal wären. ASA mit einem nicht einheitlichen Zugriffsnetzwerk wäre eine Option, um die Kosten- und Funktionsvorteile von ASA zu nutzen, ohne dass sich eine Beeinträchtigung des Zugriffs auf die standortübergreifende Latenz ergeben würde.

ASA mit SM-AS in einer Konfiguration mit niedriger Latenz bietet zwei interessante Vorteile. Zunächst verdoppelt es die Performance bei jedem einzelnen Host *, da IO von doppelt so vielen Controllern mit doppelt so vielen Pfaden gewartet werden kann. Zweitens bietet er in einer Umgebung mit einem einzigen Standort eine extreme Verfügbarkeit, da ein komplettes Storage-System ohne Unterbrechung des Host-Zugriffs verloren gehen könnte.

Uneinheitlicher Zugriff

Uneinheitliches Netzwerk durch Zugriff bedeutet, dass jeder Host nur Zugriff auf Ports im lokalen Storage-System hat. Das SAN wird nicht über Standorte (oder Ausfall-Domains am selben Standort) erweitert.



Active/Optimized Path

Der Hauptvorteil dieses Ansatzes ist die SAN-Einfachheit – Sie müssen kein SAN mehr über das Netzwerk erweitern. Einige Kunden verfügen nicht über eine Konnektivität mit niedriger Latenz zwischen den Standorten

und haben nicht die Infrastruktur, um den FC SAN-Datenverkehr über ein standortverbundenes Netzwerk zu Tunneln.

Der Nachteil eines uneinheitlichen Zugriffs besteht darin, dass bestimmte Ausfallszenarien, einschließlich des Verlusts der Replikationsverbindung, dazu führen, dass einige Hosts den Zugriff auf den Speicher verlieren. Applikationen, die als einzelne Instanzen ausgeführt werden, wie z. B. eine Datenbank ohne Cluster, die grundsätzlich nur auf einem einzelnen Host bei einem beliebigen Mount ausgeführt wird, würden ausfallen, wenn die lokale Storage-Konnektivität verloren geht. Die Daten bleiben zwar weiterhin geschützt, aber der Datenbankserver würde nicht mehr darauf zugreifen können. Es müsste an einem Remote-Standort neu gestartet werden, vorzugsweise durch einen automatisierten Prozess. VMware HA kann beispielsweise eine heruntergefahrenen Pfade auf einem Server erkennen und eine VM auf einem anderen Server neu starten, auf dem Pfade verfügbar sind.

Im Gegensatz dazu kann eine Cluster-Anwendung wie Oracle RAC einen Service bereitstellen, der gleichzeitig an zwei verschiedenen Standorten verfügbar ist. Der Verlust eines Standorts bedeutet nicht, dass der Anwendungsdienst als Ganzes verloren geht. Instanzen sind nach wie vor verfügbar und werden am verbleibenden Standort ausgeführt.

In vielen Fällen wäre die zusätzliche Latenz, wenn eine Applikation, die auf den Storage über eine Site-to-Site-Verbindung zugreift, nicht akzeptabel. Dies bedeutet, dass die verbesserte Verfügbarkeit von einheitlichem Netzwerk minimal ist, da der Verlust von Speicher an einem Standort dazu führen würde, dass die Dienste auf diesem ausgefallenen Standort sowieso heruntergefahren werden müssen.



Es gibt redundante Pfade durch den lokalen Cluster, die aus Gründen der Einfachheit nicht auf diesen Diagrammen angezeigt werden. ONTAP Storage-Systeme sind HA selbst, daher sollte ein Controller-Ausfall nicht zu einem Standortausfall führen. Es sollte lediglich zu einer Änderung führen, in der lokale Pfade auf dem betroffenen Standort verwendet werden.

Oracle Konfigurationen

Überblick

Die Verwendung von SnapMirror Active Sync trägt nicht notwendigerweise zur Ergänzung oder Änderung von Best Practices für den Betrieb einer Datenbank bei.

Die beste Architektur hängt von den geschäftlichen Anforderungen ab. Wenn das Ziel zum Beispiel ist, RPO=0 Schutz gegen Datenverlust zu haben, aber das RTO entspannt ist, dann kann die Verwendung von Oracle Single Instance Datenbanken und die Replikation der LUNs mit SM-AS ausreichen und auch preisgünstiger von einem Oracle Lizenzierungs-Standard sein. Ein Ausfall des Remote-Standorts würde den Betrieb nicht unterbrechen, und der Verlust des primären Standorts würde zu LUNs am noch intakten Standort führen, die online und einsatzbereit sind.

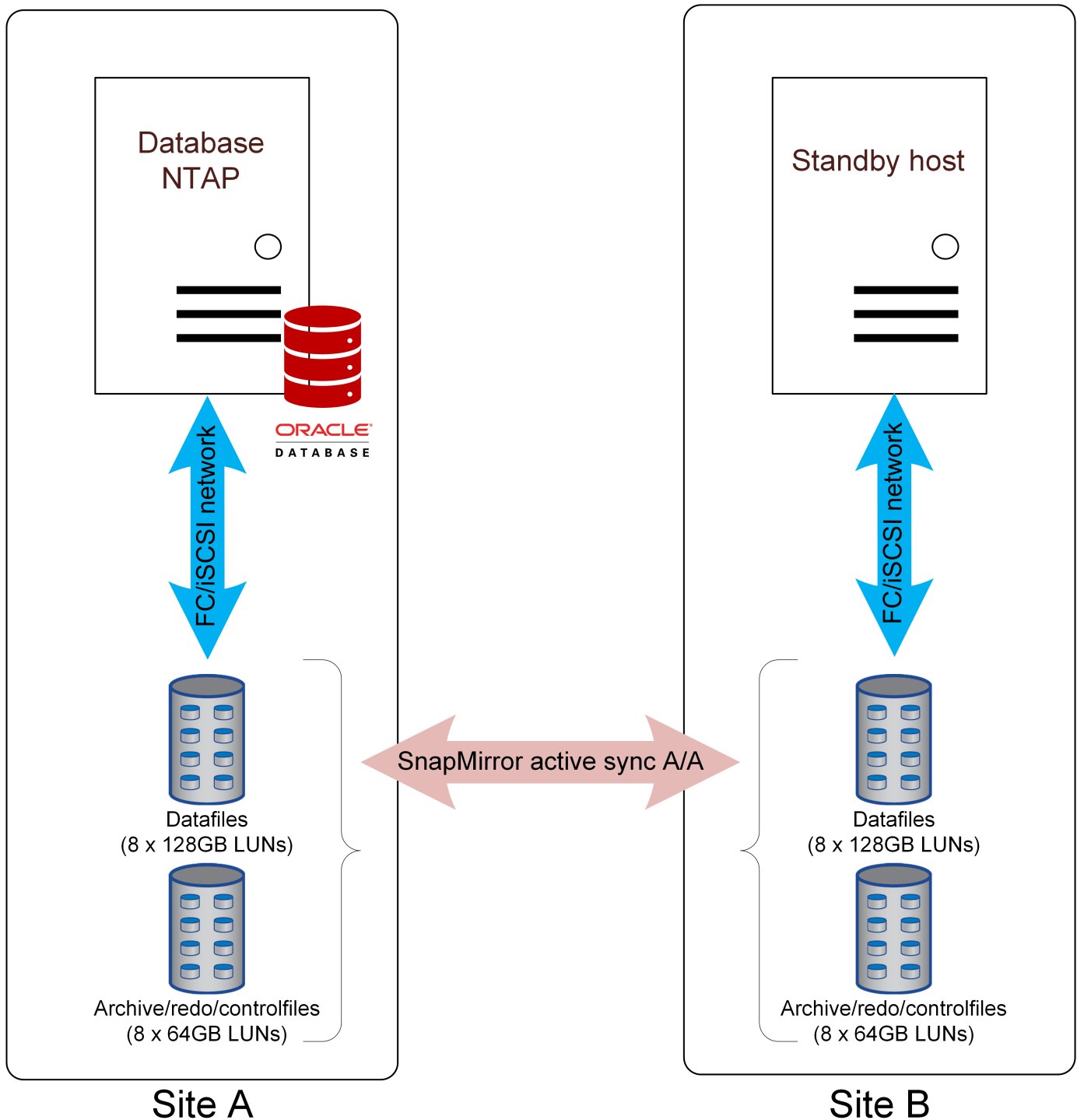
Bei einer strikteren RTO würde die grundlegende aktiv/Passiv-Automatisierung über Skripte oder Clusterware wie Pacemaker oder Ansible die Failover-Zeit verbessern. Beispielsweise könnte VMware HA konfiguriert werden, um den VM-Ausfall am primären Standort zu erkennen und die VM am Remote-Standort zu aktivieren.

Für ein extrem schnelles Failover konnte Oracle RAC über alle Standorte hinweg implementiert werden. Die RTO wäre im Grunde null, da die Datenbank jederzeit online und auf beiden Standorten verfügbar wäre.

Oracle Single Instance

Die unten erläuterten Beispiele zeigen einige der zahlreichen Optionen für die

Bereitstellung von Oracle Single-Instance-Datenbanken mit SnapMirror Active Sync-Replikation.



Failover mit einem vorkonfigurierten Betriebssystem

SnapMirror Active Sync liefert eine synchrone Kopie der Daten am Disaster Recovery-Standort. Für die Verfügbarkeit dieser Daten sind jedoch ein Betriebssystem und die zugehörigen Applikationen erforderlich. Eine grundlegende Automatisierung kann die Failover-Zeit der gesamten Umgebung deutlich verbessern. Clusterware Produkte wie Pacemaker werden häufig eingesetzt, um über die Standorte hinweg einen Cluster zu erstellen, in vielen Fällen ist der Failover-Prozess mit einfachen Skripten angesteuert.

Wenn die primären Knoten verloren gehen, stellt die Clusterware (oder Skripte) die Datenbanken am alternativen Standort online. Eine Option besteht darin, Standby-Server zu erstellen, die für die SAN-Ressourcen, aus denen die Datenbank besteht, vorkonfiguriert sind. Wenn der primäre Standort ausfällt, führt die Clusterware- oder skriptbasierte Alternative eine Abfolge von Aktionen durch, die der folgenden ähneln:

1. Fehler am primären Standort erkennen
2. Führen Sie eine Erkennung von FC- oder iSCSI-LUNs durch
3. Mounten von Dateisystemen und/oder Mounten von ASM-Datenträgergruppen
4. Die Datenbank wird gestartet

Die primäre Anforderung dieses Ansatzes ist ein Betriebssystem, das am Remote Standort ausgeführt wird. Sie muss mit Oracle-Binärdateien vorkonfiguriert sein, was auch bedeutet, dass Aufgaben wie das Patching von Oracle am primären Standort und am Standby-Standort durchgeführt werden müssen. Alternativ können die Oracle Binärdateien auf den Remote-Standort gespiegelt und gemountet werden, wenn ein Notfall deklariert wird.

Die eigentliche Aktivierung ist einfach. Befehle wie die LUN-Erkennung erfordern nur einige wenige Befehle pro FC-Port. Das Mounten von Dateisystemen ist nichts anderes als ein `mount` Befehl, und sowohl Datenbanken als auch ASM können über die CLI mit einem einzigen Befehl gestartet und gestoppt werden.

Failover mit einem virtualisierten Betriebssystem

Der Failover von Datenbankumgebungen kann auf das Betriebssystem selbst erweitert werden. In der Theorie kann dieses Failover mit Boot-LUNs durchgeführt werden, meistens erfolgt es jedoch mit einem virtualisierten Betriebssystem. Das Verfahren ähnelt den folgenden Schritten:

1. Fehler am primären Standort erkennen
2. Mounten der Datenspeicher, die die virtuellen Maschinen des Datenbankservers hosten
3. Starten der virtuellen Maschinen
4. Manuelles Starten von Datenbanken oder Konfigurieren der virtuellen Maschinen, um die Datenbanken automatisch zu starten.

Beispielsweise kann ein ESX Cluster mehrere Standorte umfassen. Bei einem Notfall können die Virtual Machines nach dem Switchover am Disaster Recovery-Standort online geschaltet werden.

Schutz vor Storage-Ausfällen

Das Diagramm oben zeigt die Verwendung von "**Uneinheitlicher Zugriff**", wo das SAN nicht über Standorte verteilt ist. Dies ist unter Umständen einfacher zu konfigurieren und ist angesichts der aktuellen SAN-Funktionen in einigen Fällen die einzige Option, was aber auch bedeutet, dass ein Ausfall des primären Storage-Systems einen Datenbankausfall bis zum Failover der Applikation zur Folge hätte.

Für zusätzliche Ausfallsicherheit könnte die Lösung mit implementiert werden "**Einheitlicher Zugriff**". Dies würde es den Anwendungen ermöglichen, den Betrieb mit den Pfaden fortzusetzen, die vom gegenüberliegenden Standort angezeigt werden.

Oracle Extended RAC

Viele Kunden optimieren ihre RTO, indem sie einen Oracle RAC Cluster über mehrere Standorte verteilen und damit eine vollständig aktiv/aktiv-Konfiguration erzielen. Das gesamte Design wird komplizierter, da es die Quorumverwaltung von Oracle RAC

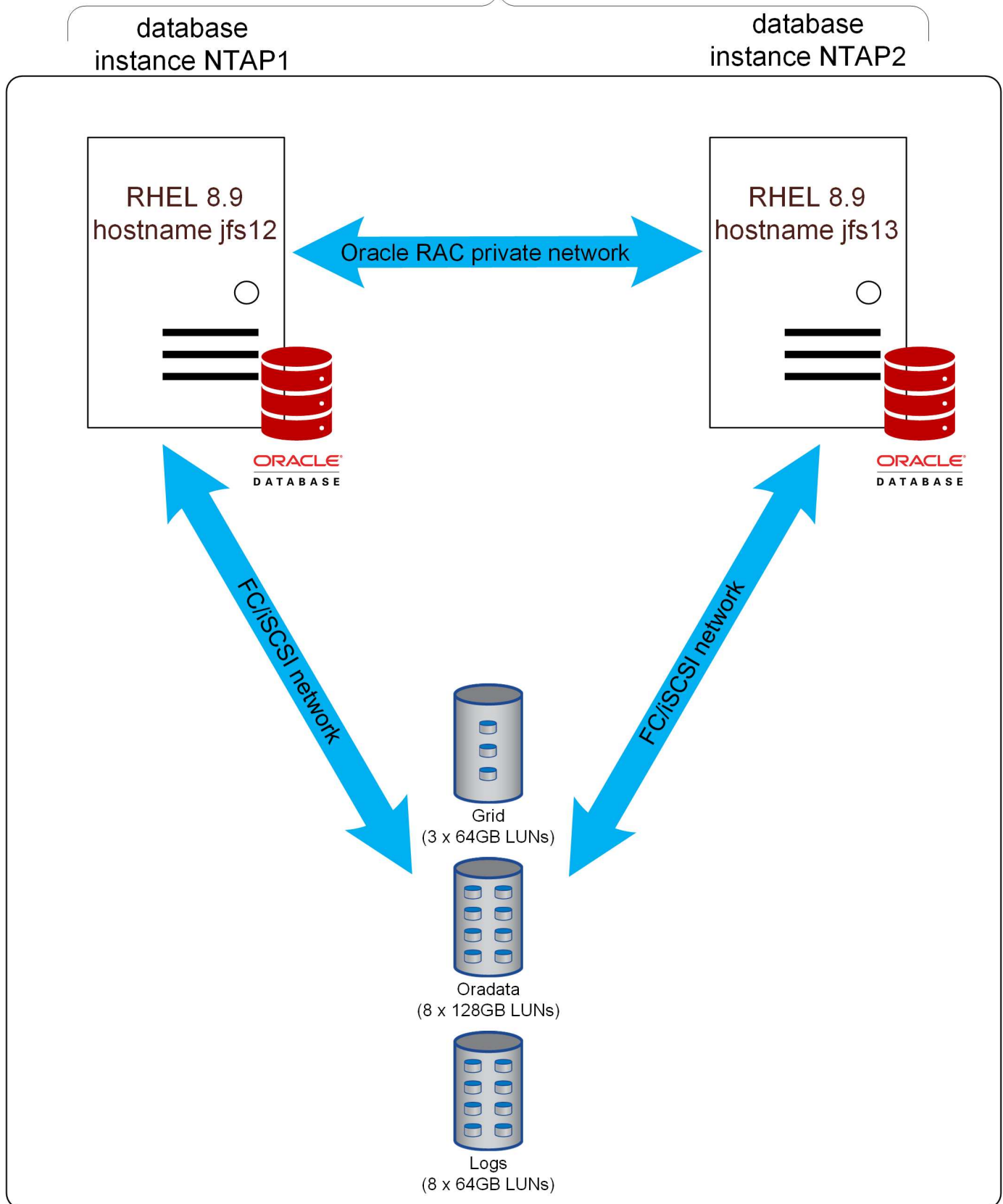
beinhalten muss.

Herkömmliche erweiterte RAC-Cluster stützten sich auf ASM-Spiegelung für Datensicherheit. Dieser Ansatz funktioniert, erfordert aber auch zahlreiche manuelle Konfigurationsschritte und führt zu einem Overhead in der Netzwerkinfrastruktur. Da hingegen die SnapMirror Active Sync die Verantwortung für die Datenreplizierung übernehmen kann, wird die Lösung erheblich vereinfacht. Vorgänge wie die Synchronisierung, die Neusynchronisierung nach Unterbrechungen, Failover und das Quorum-Management sind einfacher. Zudem muss das SAN nicht auf mehrere Standorte verteilt werden, was SAN-Design und -Management vereinfacht.

Replizierung

Sie sind für das Verständnis der RAC-Funktionalität auf SnapMirror Active Sync von zentraler Bedeutung, wenn Storage als einheitlicher Satz von LUNs angezeigt wird, die auf gespiegelter Storage gehostet werden. Beispiel:

Database NTAP



Es gibt keine primäre Kopie oder gespiegelte Kopie. Logisch gesehen, es gibt nur eine einzige Kopie jeder LUN, und diese LUN ist auf SAN-Pfaden verfügbar, die sich auf zwei verschiedenen Storage-Systemen befinden. Aus Host-Sicht gibt es keine Storage-Failover, sondern Pfadänderungen. Verschiedene

Fehlerereignisse können zum Verlust bestimmter Pfade zum LUN führen, während andere Pfade online bleiben. SnapMirror Active Sync stellt sicher, dass über alle Betriebspfade hinweg dieselben Daten verfügbar sind.

Storage-Konfiguration

In dieser Beispielkonfiguration sind die ASM-Festplatten so konfiguriert, wie sie in jeder RAC-Konfiguration mit einem einzigen Standort auf Enterprise Storage vorhanden wären. Da das Speichersystem Datenschutz bietet, würde ASM externe Redundanz verwendet werden.

Einheitlicher oder uninformatierten Zugriff

Die wichtigste Überlegung bei Oracle RAC on SnapMirror Active Sync ist, ob ein einheitlicher oder nicht einheitlicher Zugriff verwendet werden soll.

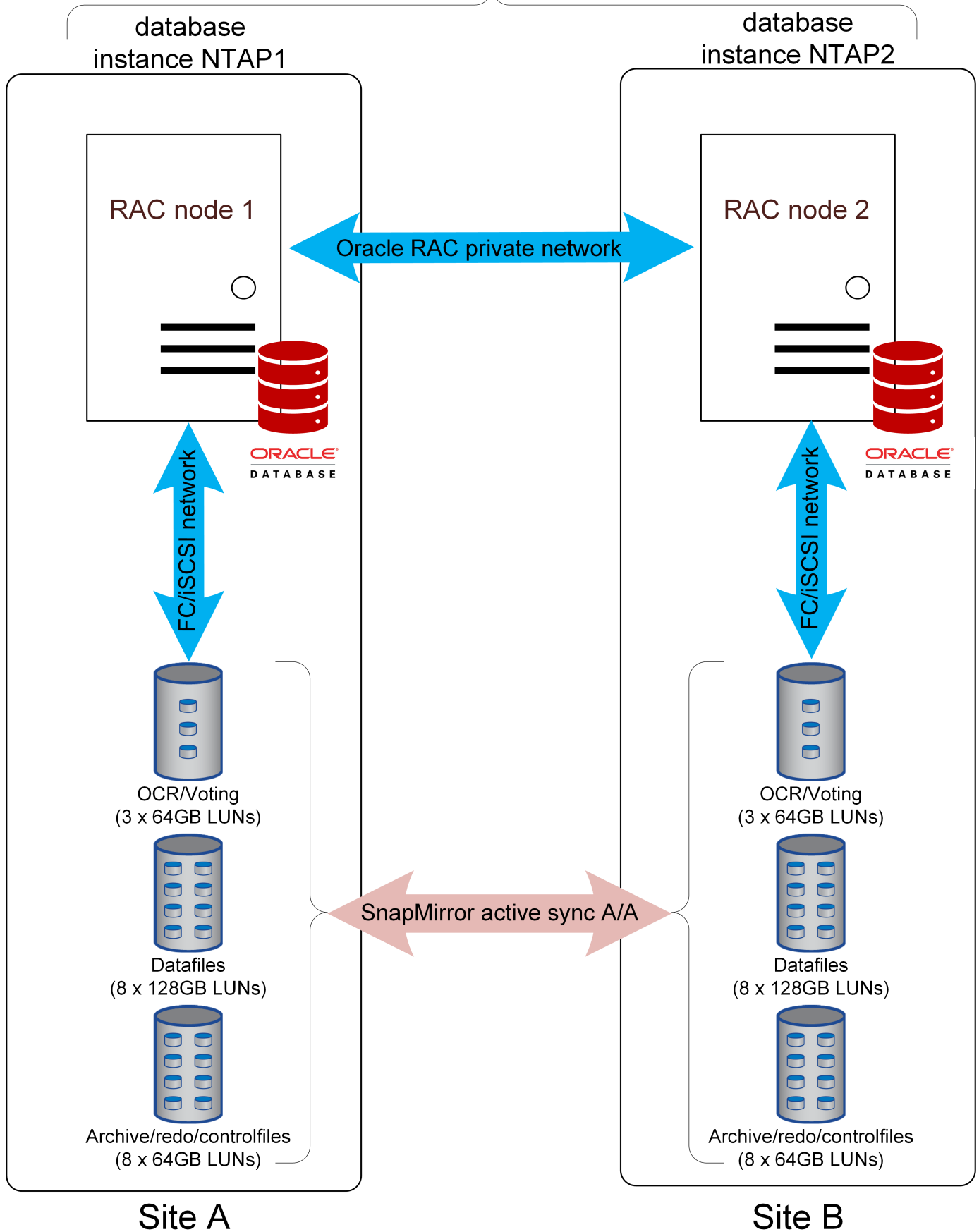
Einheitlicher Zugriff bedeutet, dass jeder Host Pfade auf beiden Clustern sehen kann. Uneinheitlicher Zugriff bedeutet, dass Hosts nur Pfade zum lokalen Cluster sehen können.

Keine Option wird ausdrücklich empfohlen oder abgeraten. Einige Kunden verfügen über Dark Fibre, um Standorte miteinander zu verbinden, andere verfügen entweder über keine solche Konnektivität oder ihre SAN-Infrastruktur unterstützt keine Long-Distance-ISL.

Uneinheitlicher Zugriff

Ein uneinheitlicher Zugriff ist aus SAN-Sicht einfacher zu konfigurieren.

Database NTAP



Der größte Nachteil des "**Uneinheitlicher Zugriff**" Ansatzes besteht darin, dass der Verlust der ONTAP-Konnektivität am Standort oder der Verlust eines Storage-Systems zum Verlust der Datenbankinstanzen an einem Standort führen kann. Dies ist natürlich nicht wünschenswert, aber es kann ein akzeptables Risiko im Austausch für eine einfachere SAN-Konfiguration sein.

Einheitlicher Zugriff

Für einen einheitlichen Zugriff muss das SAN standortübergreifend erweitert werden. Der Hauptvorteil besteht darin, dass der Verlust eines Storage-Systems nicht zum Verlust einer Datenbankinstanz führt. Stattdessen würde dies zu einer Änderung des Multipathing führen, in der Pfade derzeit verwendet werden.

Es gibt mehrere Möglichkeiten, uneinheitlichen Zugriff zu konfigurieren.

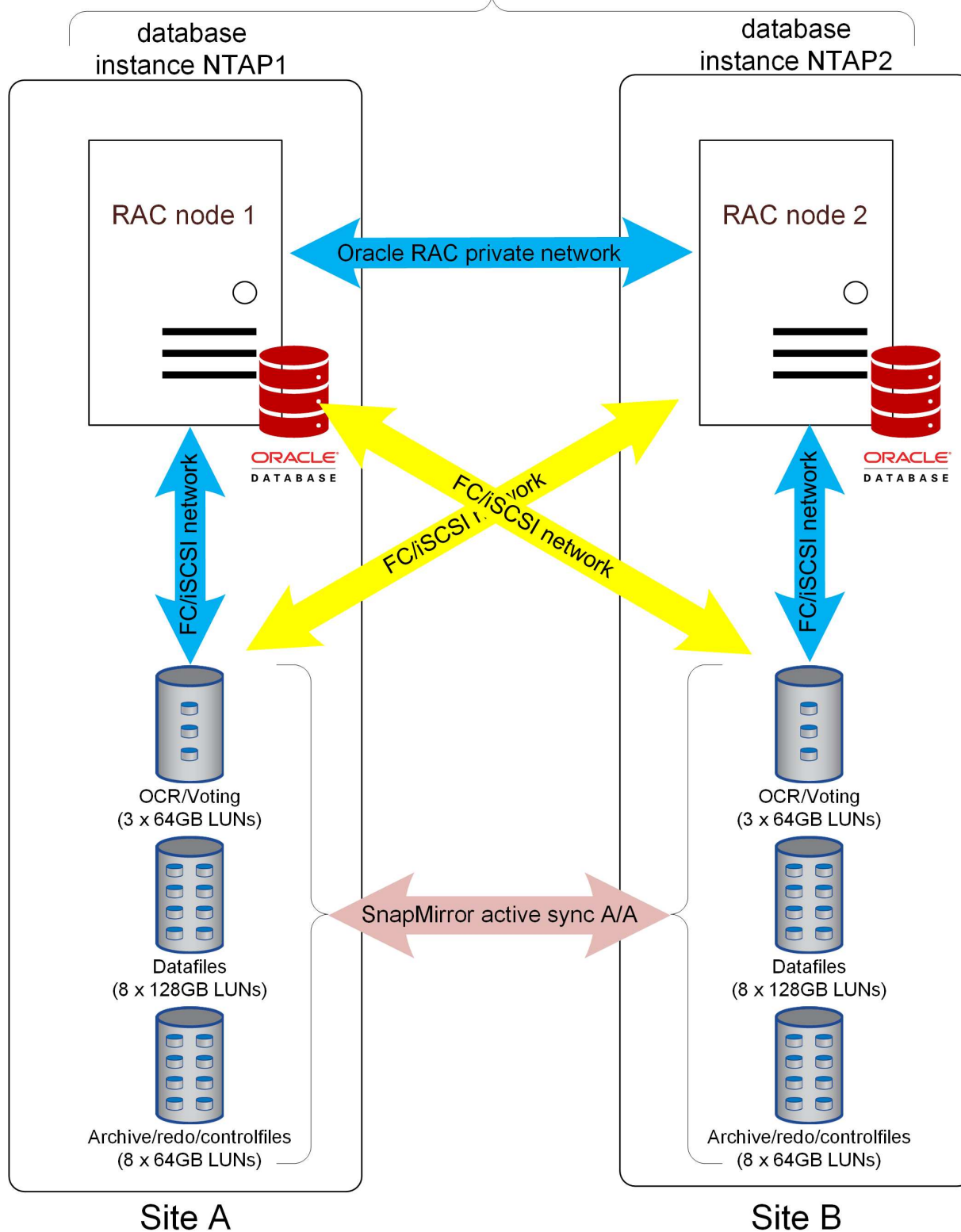


In den Diagrammen unten sind auch aktive, aber nicht optimierte Pfade vorhanden, die bei einfachen Controller-Ausfällen verwendet werden würden. Diese Pfade werden jedoch nicht im Interesse der Vereinfachung der Diagramme angezeigt.

AFF mit Annäherungseinstellungen

Bei erheblichen Latenzzeiten zwischen den Standorten können AFF Systeme mit Host-Näherungseinstellungen konfiguriert werden. So kann jedes Speichersystem erkennen, welche Hosts lokal und welche Remote sind, und Pfadprioritäten entsprechend zuweisen.

Database NTAP



Active/Optimized Path

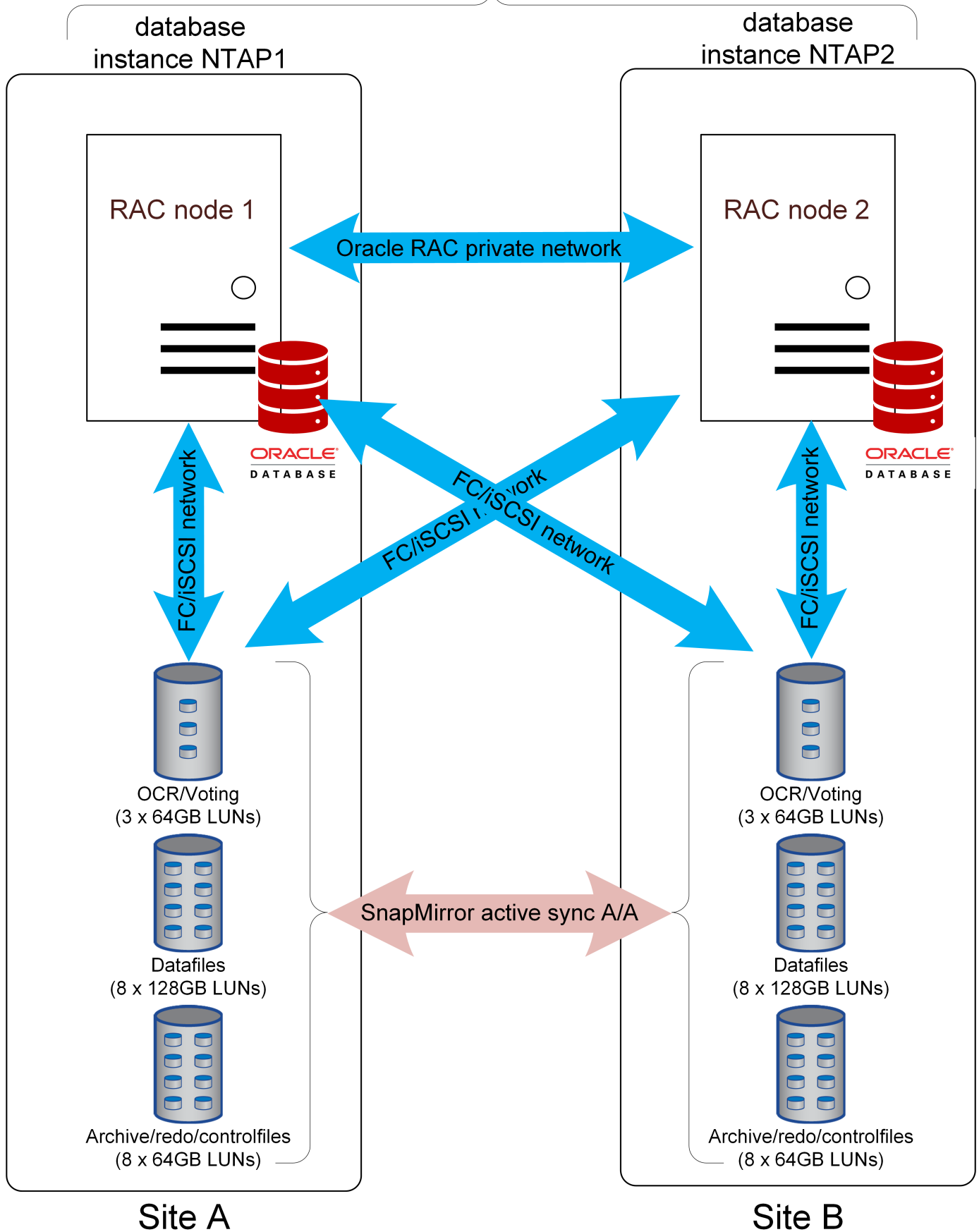
Active Path

Im normalen Betrieb würde jede Oracle-Instanz bevorzugt die lokalen aktiven/optimierten Pfade verwenden. Folglich werden alle Lesezugriffe von der lokalen Kopie der Blöcke bedient. So wird eine möglichst geringe Latenz erzielt. Schreib-I/O wird ähnlich über Pfade zum lokalen Controller gesendet. Die I/O muss noch repliziert werden, bevor sie bestätigt werden kann, und somit würde die zusätzliche Latenz beim Überqueren des Site-to-Site-Netzwerks nach wie vor entstehen. Dies kann in einer Lösung zur synchronen Replizierung jedoch nicht vermieden werden.

ASA / AFF ohne Näherungseinstellungen

Falls keine nennenswerte Latenz zwischen den Standorten erforderlich ist, können AFF Systeme ohne Host-Näherungseinstellungen konfiguriert oder ASA verwendet werden.

Database NTAP



Jeder Host kann alle Betriebspfade auf beiden Storage-Systemen verwenden. Dies verbessert potenziell die Performance erheblich, da jeder Host das Performance-Potenzial von zwei, nicht nur einem Cluster, nutzen kann.

Mit ASA gelten nicht nur alle Pfade zu beiden Clustern als aktiv und optimiert, sondern auch die Pfade auf den Partner-Controllern wären aktiv. Das Ergebnis wären ständig All-aktiv-SAN-Pfade auf dem gesamten Cluster.



ASA-Systeme können auch in einer uneinheitlichen Zugriffskonfiguration verwendet werden. Da keine standortübergreifenden Pfade vorhanden sind, würde die Performance nicht durch I/O über den ISL beeinträchtigt.

RAC Tiebreaker

Während Extended RAC mit SnapMirror Active Sync eine symmetrische Architektur in Bezug auf IO ist, gibt es eine Ausnahme, die mit Split-Brain-Management verbunden ist.

Was passiert, wenn die Replikationsverbindung verloren geht und keiner der Standorte über ein Quorum verfügt? Was soll geschehen? Diese Frage bezieht sich sowohl auf das Oracle RAC- als auch auf das ONTAP-Verhalten. Wenn Änderungen nicht standortübergreifend repliziert werden können und Sie den Betrieb wieder aufnehmen möchten, muss einer der Standorte überleben und der andere Standort muss nicht mehr verfügbar sein.

Das **"ONTAP Mediator"** löst diese Anforderung auf ONTAP-Ebene. Es gibt mehrere Optionen für RAC Tiebreaking.

Oracle Tiebreakers

Die beste Methode zur Verwaltung von Split-Brain Oracle RAC-Risiken ist die Verwendung einer ungeraden Anzahl von RAC-Knoten, vorzugsweise unter Verwendung eines Tiebreaker am dritten Standort. Wenn ein dritter Standort nicht verfügbar ist, könnte die Tiebreaker Instanz auf einem Standort der beiden Standorte platziert werden und somit einen bevorzugten Survivor-Standort darstellen.

Oracle und `css_critical`

Bei einer geraden Anzahl von Knoten ist das standardmäßige Oracle RAC-Verhalten, dass einer der Knoten im Cluster als wichtiger angesehen wird als die anderen Knoten. Der Standort mit diesem Knoten mit höherer Priorität übersteht die Standortisolierung, während die Knoten am anderen Standort entfernt werden. Die Priorisierung basiert auf mehreren Faktoren, aber Sie können dieses Verhalten auch über die Einstellung `css_critical` steuern.

In der **"Beispiel"** Architektur sind die Hostnamen für die RAC-Knoten `jfs12` und `jfs13`. Die aktuellen Einstellungen für `css_critical` sind wie folgt:

```
[root@jfs12 ~]# /grid/bin/crsctl get server css_critical
CRS-5092: Current value of the server attribute CSS_CRITICAL is no.

[root@jfs13 trace]# /grid/bin/crsctl get server css_critical
CRS-5092: Current value of the server attribute CSS_CRITICAL is no.
```

Wenn der Standort mit `jfs12` der bevorzugte Standort sein soll, ändern Sie diesen Wert für einen Knoten an Standort A in Ja, und starten Sie die Dienste neu.


```
[root@jfs12 ~]# /grid/bin/crsctl set server css_critical yes
CRS-4416: Server attribute 'CSS_CRITICAL' successfully changed. Restart
Oracle High Availability Services for new value to take effect.

[root@jfs12 ~]# /grid/bin/crsctl stop crs
CRS-2791: Starting shutdown of Oracle High Availability Services-managed
resources on 'jfs12'
CRS-2673: Attempting to stop 'ora.crsd' on 'jfs12'
CRS-2790: Starting shutdown of Cluster Ready Services-managed resources on
server 'jfs12'
CRS-2673: Attempting to stop 'ora.ntap.ntappdb1.pdb' on 'jfs12'
...
CRS-2673: Attempting to stop 'ora.gipcd' on 'jfs12'
CRS-2677: Stop of 'ora.gipcd' on 'jfs12' succeeded
CRS-2793: Shutdown of Oracle High Availability Services-managed resources
on 'jfs12' has completed
CRS-4133: Oracle High Availability Services has been stopped.

[root@jfs12 ~]# /grid/bin/crsctl start crs
CRS-4123: Oracle High Availability Services has been started.
```

Ausfallszenarien

Überblick

Die Planung einer vollständigen Applikationsarchitektur für die aktive Synchronisierung von SnapMirror erfordert ein Verständnis dafür, wie SM-AS in verschiedenen geplanten und ungeplanten Failover-Szenarien reagiert.

In den folgenden Beispielen wird davon ausgegangen, dass Standort A als bevorzugter Standort konfiguriert ist.

Verlust der Replikationskonnektivität

Wenn die SM-AS-Replikation unterbrochen wird, kann die Schreib-I/O nicht abgeschlossen werden, da ein Cluster Änderungen nicht auf den anderen Standort replizieren kann.

Standort A (bevorzugte Website)

Das Ergebnis eines Ausfalls der Replikationsverbindung auf dem bevorzugten Standort ist eine ca. 15-Sekunden-Pause bei der Schreib-I/O-Verarbeitung, da ONTAP erneut replizierte Schreibvorgänge versucht, bevor festgestellt wird, dass die Replikationsverbindung wirklich nicht erreichbar ist. Nach 15 Sekunden wird die I/O-Verarbeitung von Lese- und Schreibzugriffen von Standort A fortgesetzt. Die SAN-Pfade ändern sich nicht, und die LUNs bleiben online.

Standort B

Da Standort B nicht der bevorzugte Standort für SnapMirror Active Sync ist, sind die LUN-Pfade nach ca. 15 Sekunden nicht mehr verfügbar.

Ausfall des Storage-Systems

Das Ergebnis eines Storage-Systemausfalls ist nahezu identisch mit dem Ergebnis des Verlusts der Replizierungsverbindung. Am überlebenden Standort sollte eine I/O-Pause von etwa 15 Sekunden stattfinden. Nach Ablauf dieses Zeitraums von 15 Sekunden wird die E/A-Vorgänge wie gewohnt an diesem Standort fortgesetzt.

Verlust des Mediators

Der Mediator hat keine direkte Kontrolle über Storage-Vorgänge. Er fungiert als alternativer Kontrollpfad zwischen Clustern. Die Lösung bietet insbesondere automatisierte Failover-Prozesse ohne Split-Brain-Szenario. Im normalen Betrieb repliziert jedes Cluster Änderungen an seinem Partner. Daher kann jedes Cluster überprüfen, ob das Partner-Cluster online ist und Daten bereitstellt. Wenn die Replikationsverbindung fehlschlägt, wird die Replikation beendet.

Der Grund für einen sicheren automatisierten Failover ist der Mediator, der darauf zurückzuführen ist, dass ein Storage-Cluster andernfalls nicht feststellen kann, ob der Ausfall einer bidirektionalen Kommunikation auf einen Netzwerkausfall oder einen tatsächlichen Storage-Ausfall zurückzuführen ist.

Der Mediator bietet jedem Cluster einen alternativen Pfad zur Überprüfung des Integrität seines Partners. Die Szenarien sind wie folgt:

- Wenn ein Cluster seinen Partner direkt kontaktieren kann, sind die Replizierungsservices betriebsbereit. Keine Aktion erforderlich.
- Wenn ein bevorzugter Standort nicht direkt mit dem Partner oder über den Mediator in Kontakt treten kann, wird davon ausgegangen, dass der Partner entweder tatsächlich nicht verfügbar ist oder isoliert wurde und seine LUN-Pfade offline geschaltet hat. Der bevorzugte Standort setzt dann den Status RPO=0 frei und setzt die Verarbeitung von Lese- und Schreib-I/O fort.
- Wenn ein nicht bevorzugter Standort seinen Partner nicht direkt kontaktieren kann, ihn aber über den Mediator kontaktieren kann, nimmt er seine Pfade offline und wartet auf die Rückkehr der Replikationsverbindung.
- Wenn ein nicht bevorzugter Standort keine direkte Kontaktaufnahme mit dem Partner oder über einen betrieblichen Mediator bietet, nimmt er an, dass der Partner entweder tatsächlich nicht verfügbar ist oder isoliert war und seine LUN-Pfade offline geschaltet hat. Der nicht bevorzugte Standort setzt dann den Status RPO=0 frei und verarbeitet sowohl Lese- als auch Schreib-I/O weiter. Er übernimmt die Rolle der Replikationsquelle und wird der neue bevorzugte Standort.

Wenn der Mediator vollständig nicht verfügbar ist:

- Wenn keine Replizierungsservices aus irgendeinem Grund verfügbar sind, beispielsweise der Ausfall des nicht bevorzugten Standorts oder des Storage-Systems, wird der bevorzugte Standort den Zustand RPO=0 freigeben und die I/O-Verarbeitung für Lese- und Schreibvorgänge wird wieder aufgenommen. Der nicht bevorzugte Standort nimmt seine Pfade offline.
- Ein Ausfall des bevorzugten Standorts führt zu einem Ausfall, da der nicht bevorzugte Standort nicht verifizieren kann, dass der gegenteilige Standort wirklich offline ist. Daher ist es für den nicht bevorzugten Standort nicht sicher, die Services wieder aufzunehmen.

Dienste werden wiederhergestellt

Wenn ein Fehler behoben wurde, wie z. B. die Wiederherstellung der Site-to-Site-Verbindung oder das Einschalten eines ausgefallenen Systems, erkennen die SnapMirror Active Sync-Endpunkte automatisch, dass eine fehlerhafte Replikationsbeziehung vorhanden ist, und versetzen sie wieder in den Zustand RPO=0. Sobald die synchrone Replizierung wiederhergestellt ist, werden die fehlerhaften Pfade wieder online geschaltet.

In vielen Fällen erkennen Cluster-Applikationen automatisch die Rückgabe ausgefallener Pfade, und diese Applikationen sind ebenfalls wieder online. In anderen Fällen ist möglicherweise ein SAN-Scan auf Host-Ebene erforderlich oder Applikationen müssen manuell wieder online geschaltet werden. Es hängt von der Anwendung und ihrer Konfiguration ab, und im Allgemeinen lassen sich solche Aufgaben leicht automatisieren. ONTAP selbst behebt selbstständig und sollte keinen Benutzereingriff erfordern, um den RPO=0-Storage-Betrieb wiederaufzunehmen.

Manueller Failover

Das Ändern des bevorzugten Standorts erfordert eine einfache Bedienung. I/O-Vorgänge werden für eine oder zwei Sekunden angehalten, da zwischen den Clustern die Berechtigung für das Replikationsverhalten wechselt, die E/A-Vorgänge sind jedoch ansonsten nicht betroffen.

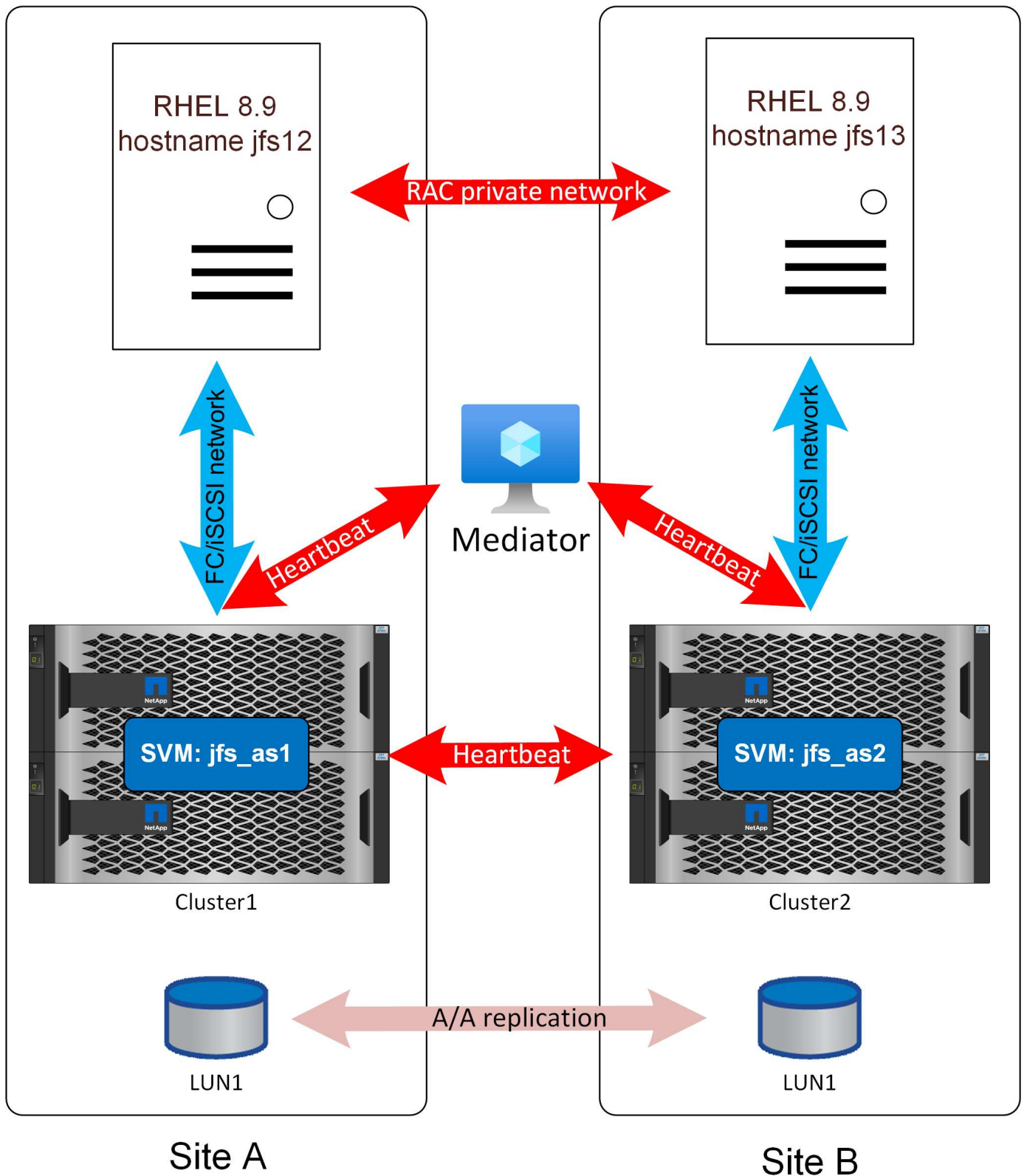
Beispielarchitektur

Die in diesen Abschnitten gezeigten detaillierten Fehlerbeispiele basieren auf der unten dargestellten Architektur.



Dies ist nur eine von vielen Optionen für Oracle Datenbanken auf SnapMirror Active Sync. Dieses Design wurde gewählt, weil es einige der komplizierteren Szenarien illustriert.

Bei diesem Design wird davon ausgegangen, dass Standort A auf der eingestellt ist "**Bevorzugter Standort**".



RAC-Verbindungsfehler

Der Verlust der Oracle RAC-Replikationsverbindung führt zu einem ähnlichen Ergebnis wie der Verlust der SnapMirror-Konnektivität, mit Ausnahme der standardmäßig kürzeren Timeouts. In den Standardeinstellungen wartet ein Oracle RAC-Knoten 200 Sekunden

nach Verlust der Speicherverbindung, bevor er entfernt wird, aber er wartet nur 30 Sekunden nach Verlust des RAC-Netzwerk-Heartbeat.

Die CRS-Meldungen ähneln denen unten. Sie können die Zeitlimitüberschreitung von 30 Sekunden sehen. Da `css_Critical` auf `jfs12` gesetzt wurde, befindet sich an Ort A, das wird die Website zu überleben und `jfs13` auf Standort B wird entfernt werden.

```
2024-09-12 10:56:44.047 [ONMD(3528)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval. If this
persists, removal of this node from cluster will occur in 6.980 seconds
2024-09-12 10:56:48.048 [ONMD(3528)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval. If this
persists, removal of this node from cluster will occur in 2.980 seconds
2024-09-12 10:56:51.031 [ONMD(3528)]CRS-1607: Node jfs13 is being evicted
in cluster incarnation 621599354; details at (:CSSNM00007:) in
/gridbase/diag/crs/jfs12/crs/trace/onmd.trc.
2024-09-12 10:56:52.390 [CRSD(6668)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:33194;', interface list of remote node 'jfs13' is
'192.168.30.2:33621;'.
2024-09-12 10:56:55.683 [ONMD(3528)]CRS-1601: CSSD Reconfiguration
complete. Active nodes are jfs12 .
2024-09-12 10:56:55.722 [CRSD(6668)]CRS-5504: Node down event reported for
node 'jfs13'.
2024-09-12 10:56:57.222 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'Generic'.
2024-09-12 10:56:57.224 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'ora.NTAP'.
```

SnapMirror-Kommunikationsfehler

Wenn der SnapMirror Active Sync Replication Link verwendet wird, kann die Schreib-I/O nicht abgeschlossen werden, da ein Cluster Änderungen nicht am anderen Standort replizieren könnte.

Standort A

Das Ergebnis eines Ausfalls einer Replikationsverbindung an Standort A ist eine ca. 15-Sekunden-Pause bei der Schreib-I/O-Verarbeitung, da ONTAP versucht, Schreibvorgänge zu replizieren, bevor es feststellt, dass die Replikationsverbindung wirklich nicht funktionsfähig ist. Nach 15 Sekunden wird das ONTAP Cluster vor Ort A mit der Lese- und Schreib-I/O-Verarbeitung fortgesetzt. Die SAN-Pfade ändern sich nicht, und die LUNs bleiben online.

Standort B

Da Standort B nicht der bevorzugte Standort für SnapMirror Active Sync ist, sind die LUN-Pfade nach ca. 15 Sekunden nicht mehr verfügbar.

Die Replikationsverbindung wurde mit dem Zeitstempel 15:19:44 geschnitten. Die erste Warnung von Oracle RAC kommt 100 Sekunden später, da sich das 200-Sekunden-Timeout (gesteuert durch den Oracle RAC Parameter disktimeout) nähert.

```
2024-09-10 15:21:24.702 [ONMD(2792)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99340 milliseconds.
2024-09-10 15:22:14.706 [ONMD(2792)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49330 milliseconds.
2024-09-10 15:22:44.708 [ONMD(2792)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19330 milliseconds.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.716 [ONMD(2792)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.731 [OCSSD(2794)]CRS-1652: Starting clean up of CRS
resources.
```

Sobald das 200-Sekunden-Zeitlimit für Abstimmdateinträger erreicht wurde, wird dieser Oracle RAC-Knoten selbst aus dem Cluster entfernt und neu gestartet.

Totaler Fehler bei der Netzwerkverbindung

Wenn die Replikationsverbindung zwischen den Standorten vollständig unterbrochen wird, werden sowohl die aktive SnapMirror-Synchronisierung als auch die Oracle RAC-Verbindung unterbrochen.

Die Split-Brain-Erkennung von Oracle RAC ist vom Heartbeat des Oracle RAC Storage abhängig. Wenn der Verlust der Site-to-Site-Konnektivität zu einem gleichzeitigen Verlust sowohl des RAC-Netzwerk-Heartbeat als auch der Speicherreplikationsdienste führt, können die RAC-Standorte weder über das RAC-Interconnect noch über die RAC-Abstimmungs-Laufwerke standortübergreifend kommunizieren. Das Ergebnis einer geraden Anzahl von Knoten kann die Entfernung beider Standorte unter den Standardeinstellungen sein. Das genaue Verhalten hängt von der Reihenfolge der Ereignisse und dem Timing des RAC-Netzwerks und der Disk-Heartbeat-Abfragen ab.

Das Risiko eines Ausfalls von 2 Standorten kann auf zwei Arten behoben werden. Zunächst kann eine **"Tiebreaker"** Konfiguration verwendet werden.

Wenn kein dritter Standort verfügbar ist, kann dieses Risiko durch Anpassung des Parameters für die

Fehlzählung im RAC-Cluster behoben werden. Unter den Standardeinstellungen beträgt das Heartbeat-Timeout des RAC-Netzwerks 30 Sekunden. Dies wird normalerweise von RAC verwendet, um fehlerhafte RAC-Knoten zu identifizieren und aus dem Cluster zu entfernen. Es hat auch eine Verbindung zum Abstimmmedium Heartbeat.

Wenn beispielsweise das Verbindungsrohr, das den Datenverkehr zwischen den Standorten für Oracle RAC und Speicherreplikationsdienste transportiert, durch einen Bagger gekürzt wird, beginnt der 30-Sekunden-Countdown für die Fehlzählung. Wenn der bevorzugte RAC-Standortknoten den Kontakt zum anderen Standort nicht innerhalb von 30 Sekunden wiederherstellen kann und er auch nicht die Abstimmdisks verwenden kann, um zu bestätigen, dass sich der entgegengesetzte Standort innerhalb desselben 30-Sekunden-Fensters befindet, werden die bevorzugten Standortknoten ebenfalls entfernt. Das Ergebnis ist ein vollständiger Ausfall der Datenbank.

Je nachdem, wann die Abfrage der Fehlzählung erfolgt, sind 30 Sekunden möglicherweise nicht genügend Zeit für die SnapMirror Active Sync, um die Zeit zu verkürzen und die Speicherung auf dem bevorzugten Standort zu ermöglichen, um die Dienste wieder aufzunehmen, bevor das 30-Sekunden-Fenster abläuft. Dieses 30-Sekunden-Fenster kann vergrößert werden.

```
[root@jfs12 ~]# /grid/bin/crsctl set css misscount 100
CRS-4684: Successful set of parameter misscount to 100 for Cluster
Synchronization Services.
```

Mit diesem Wert kann das Speichersystem am bevorzugten Standort den Betrieb wieder aufnehmen, bevor das Timeout für die Fehlzählung abläuft. Das Ergebnis ist eine Entfernung nur der Knoten am Standort, an dem die LUN-Pfade entfernt wurden. Beispiel unten:

```
2024-09-12 09:50:59.352 [ONMD(681360)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval. If this
persists, removal of this node from cluster will occur in 49.570 seconds
2024-09-12 09:51:10.082 [CRSD(682669)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:46039;', interface list of remote node 'jfs13' is
'192.168.30.2:42037;'.
2024-09-12 09:51:24.356 [ONMD(681360)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval. If this
persists, removal of this node from cluster will occur in 24.560 seconds
2024-09-12 09:51:39.359 [ONMD(681360)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval. If this
persists, removal of this node from cluster will occur in 9.560 seconds
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8011: reboot advisory message
from host: jfs13, component: cssagent, with time stamp: L-2024-09-12-
09:51:47.451
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8013: reboot advisory message
text: oracssdagent is about to reboot this node due to unknown reason as
it did not receive local heartbeats for 10470 ms amount of time
2024-09-12 09:51:48.925 [ONMD(681360)]CRS-1632: Node jfs13 is being
removed from the cluster in cluster incarnation 621596607
```

Der Oracle Support rät dringend davon ab, die Parameter „Fehlstellen“ oder „Disktimeout“ zu ändern, um Konfigurationsprobleme zu lösen. Eine Änderung dieser Parameter kann jedoch in vielen Fällen gerechtfertigt und unvermeidbar sein, einschließlich Konfigurationen für SAN-Bootting, virtualisierte Konfigurationen und Speicherreplikation. Wenn Sie beispielsweise Stabilitätsprobleme mit einem SAN- oder IP-Netzwerk hatten, das zu RAC-Räumungen führte, sollten Sie das zugrunde liegende Problem beheben und die Werte des Misscount- oder Disktimeout nicht aufladen. Durch das Ändern von Timeouts zur Behebung von Konfigurationsfehlern wird ein Problem maskiert und kein Problem gelöst. Die Änderung dieser Parameter zur ordnungsgemäßen Konfiguration einer RAC-Umgebung basierend auf Designaspekten der zugrunde liegenden Infrastruktur unterscheidet sich und entspricht den Oracle-Support-Anweisungen. Bei SAN-Bootvorgang ist es üblich, Fehlstellen bis zu 200 anzupassen, um Disktimeout zu entsprechen. Weitere Informationen finden Sie unter ["Dieser Link"](#).

Standortausfall

Das Ergebnis eines Storage-System- oder Standortausfalls ist nahezu identisch mit dem Ergebnis des Verlusts der Replizierungsverbindung. Am verbleibenden Standort sollte eine I/O-Pause von etwa 15 Sekunden bei Schreibvorgängen stattfinden. Nach Ablauf dieses Zeitraums von 15 Sekunden wird die E/A-Vorgänge wie gewohnt an diesem Standort fortgesetzt.

Wenn nur das Speichersystem betroffen war, gehen die Speicherdienste des Oracle RAC-Knotens am ausgefallenen Standort verloren und führen vor der Entfernung und dem anschließenden Neustart denselben Countdown für die 200-Sekunden-Zeitüberschreitung für die Festplatte ein.


```

2024-09-11 13:44:38.613 [ONMD(3629)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99750 milliseconds.
2024-09-11 13:44:51.202 [ORAAGENT(5437)]CRS-5011: Check of resource "NTAP"
failed: details at "(:CLSN00007:)" in
"/gridbase/diag/crs/jfs13/crs/trace/crsd_oraagent_oracle.trc"
2024-09-11 13:44:51.798 [ORAAGENT(75914)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 75914
2024-09-11 13:45:28.626 [ONMD(3629)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49730 milliseconds.
2024-09-11 13:45:33.339 [ORAAGENT(76328)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 76328
2024-09-11 13:45:58.629 [ONMD(3629)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19730 milliseconds.
2024-09-11 13:46:18.630 [ONMD(3629)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-11 13:46:18.631 [ONMD(3629)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.638 [ONMD(3629)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.651 [OCSSD(3631)]CRS-1652: Starting clean up of CRS
resources.

```

Der SAN-Pfadstatus auf dem RAC-Knoten, der die Speicherdienste verloren hat, sieht wie folgt aus:

```

oradata7 (3600a0980383041334a3f55676c697347) dm-20 NETAPP,LUN C-Mode
size=128G features='3 queue_if_no_path pg_init_retries 50' hwhandler='1
alua' wp=rw
|-+- policy='service-time 0' prio=0 status=enabled
|  - 34:0:0:18 sdam 66:96  failed faulty running
`-+- policy='service-time 0' prio=0 status=enabled
   - 33:0:0:18 sdaj 66:48  failed faulty running

```

Der linux-Host hat den Verlust der Pfade viel schneller als 200 Sekunden erkannt, aber aus Sicht der Datenbank werden die Clientverbindungen zum Host auf dem ausgefallenen Standort unter den standardmäßigen Oracle RAC-Einstellungen weiterhin 200 Sekunden lang eingefroren. Die vollständigen Datenbankvorgänge werden erst nach Abschluss der Entfernung fortgesetzt.

In der Zwischenzeit zeichnet der Oracle RAC-Knoten am gegenüberliegenden Standort den Verlust des anderen RAC-Knotens auf. Ansonsten funktioniert es wie gewohnt.

```
2024-09-11 13:46:34.152 [ONMD(3547)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval. If this
persists, removal of this node from cluster will occur in 14.020 seconds
2024-09-11 13:46:41.154 [ONMD(3547)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval. If this
persists, removal of this node from cluster will occur in 7.010 seconds
2024-09-11 13:46:46.155 [ONMD(3547)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval. If this
persists, removal of this node from cluster will occur in 2.010 seconds
2024-09-11 13:46:46.470 [OHASD(1705)]CRS-8011: reboot advisory message
from host: jfs13, component: cssmonit, with time stamp: L-2024-09-11-
13:46:46.404
2024-09-11 13:46:46.471 [OHASD(1705)]CRS-8013: reboot advisory message
text: At this point node has lost voting file majority access and
oracssdmonitor is rebooting the node due to unknown reason as it did not
receive local hearbeats for 28180 ms amount of time
2024-09-11 13:46:48.173 [ONMD(3547)]CRS-1632: Node jfs13 is being removed
from the cluster in cluster incarnation 621516934
```

Fehler beim Mediator

Der Mediator hat keine direkte Kontrolle über Storage-Vorgänge. Er fungiert als alternativer Kontrollpfad zwischen Clustern. Die Lösung bietet insbesondere automatisierte Failover-Prozesse ohne Split-Brain-Szenario.

Im normalen Betrieb repliziert jedes Cluster Änderungen an seinem Partner. Daher kann jedes Cluster überprüfen, ob das Partner-Cluster online ist und Daten bereitstellt. Wenn die Replikationsverbindung fehlschlägt, wird die Replikation beendet.

Der Grund für den sicheren automatisierten Betrieb ist ein Mediator, der andernfalls nicht feststellen kann, ob der Ausfall einer bidirektionalen Kommunikation auf einen Netzwerkausfall oder einen tatsächlichen Storage-Ausfall zurückzuführen ist.

Der Mediator bietet jedem Cluster einen alternativen Pfad zur Überprüfung des Integrität seines Partners. Die Szenarien sind wie folgt:

- Wenn ein Cluster seinen Partner direkt kontaktieren kann, sind die Replizierungsservices betriebsbereit. Keine Aktion erforderlich.
- Wenn ein bevorzugter Standort nicht direkt mit dem Partner oder über den Mediator in Kontakt treten kann, wird davon ausgegangen, dass der Partner entweder tatsächlich nicht verfügbar ist oder isoliert wurde und seine LUN-Pfade offline geschaltet hat. Der bevorzugte Standort setzt dann den Status RPO=0 frei und setzt die Verarbeitung von Lese- und Schreib-I/O fort.
- Wenn ein nicht bevorzugter Standort seinen Partner nicht direkt kontaktieren kann, ihn aber über den Mediator kontaktieren kann, nimmt er seine Pfade offline und wartet auf die Rückkehr der Replikationsverbindung.

- Wenn ein nicht bevorzugter Standort keine direkte Kontaktaufnahme mit dem Partner oder über einen betrieblichen Mediator bietet, nimmt er an, dass der Partner entweder tatsächlich nicht verfügbar ist oder isoliert war und seine LUN-Pfade offline geschaltet hat. Der nicht bevorzugte Standort setzt dann den Status RPO=0 frei und verarbeitet sowohl Lese- als auch Schreib-I/O weiter. Er übernimmt die Rolle der Replikationsquelle und wird der neue bevorzugte Standort.

Wenn der Mediator vollständig nicht verfügbar ist:

- Ein Ausfall der Replikationsdienste führt aus irgendeinem Grund dazu, dass der bevorzugte Standort den Zustand RPO=0 freigibt und die Lese- und Schreib-I/O-Verarbeitung wieder aufgenommen wird. Der nicht bevorzugte Standort nimmt seine Pfade offline.
- Ein Ausfall des bevorzugten Standorts führt zu einem Ausfall, da der nicht bevorzugte Standort nicht verifizieren kann, dass der gegenteilige Standort wirklich offline ist. Daher ist es für den nicht bevorzugten Standort nicht sicher, die Services wieder aufzunehmen.

Servicewiederherstellung

SnapMirror bietet Selbstreparatur. SnapMirror Active Sync erkennt automatisch eine fehlerhafte Replikationsbeziehung und versetzt sie zurück in den Zustand RPO=0. Sobald die synchrone Replikation wiederhergestellt ist, werden die Pfade wieder online geschaltet.

In vielen Fällen erkennen Cluster-Applikationen automatisch die Rückgabe ausgefallener Pfade, und diese Applikationen sind ebenfalls wieder online. In anderen Fällen ist möglicherweise ein SAN-Scan auf Host-Ebene erforderlich oder Applikationen müssen manuell wieder online geschaltet werden.

Es hängt von der Anwendung und ihrer Konfiguration ab, und im Allgemeinen können solche Aufgaben leicht automatisiert werden. Die SnapMirror Active Sync Software selbst wird automatisch behoben und sollte nach der Wiederherstellung der Stromversorgung und Konnektivität keinen Benutzereingriff erfordern, um die RPO=0-Speichervorgänge wiederaufzunehmen.

Manueller Failover

Der Begriff „Failover“ bezieht sich nicht auf die Richtung der Replizierung mit SnapMirror Active Sync, da es sich um eine bidirektionale Replizierungstechnologie handelt. Stattdessen bezieht sich „Failover“ darauf, welches Speichersystem bei einem Ausfall der bevorzugte Standort ist.

Möglicherweise möchten Sie beispielsweise ein Failover ausführen, um den bevorzugten Standort zu ändern, bevor Sie einen Standort zu Wartungszwecken herunterfahren oder bevor Sie einen DR-Test durchführen.

Das Ändern des bevorzugten Standorts erfordert eine einfache Bedienung. I/O-Vorgänge werden für eine oder zwei Sekunden angehalten, da zwischen den Clustern die Berechtigung für das Replikationsverhalten wechselt, die E/A-Vorgänge sind jedoch ansonsten nicht betroffen.

GUI-Beispiel:

Relationships

Local destinations

Local sources

Search Download Show/hide Filter

Source	Destination	Policy type
jfs_as1:/cg/jfsAA	jfs_as2:/cg/jfsAA	Synchronous
<div>Edit</div> <div>Update</div> <div>Delete</div> <div>Failover</div>		

Beispiel für eine Rückänderung über die CLI:

```
Cluster2::> snapmirror failover start -destination-path jfs_as2:/cg/jfsAA
[Job 9575] Job is queued: SnapMirror failover for destination
"jfs_as2:/cg/jfsAA".
```

```
Cluster2::> snapmirror failover show
```

Source Path	Destination Path	Type	Status	start-time	end-time	Error Reason
jfs_as1:/cg/jfsAA	jfs_as2:/cg/jfsAA	planned	completed	9/11/2024 09:29:22	9/11/2024 09:29:32	

The new destination path can be verified as follows:

```
Cluster1::> snapmirror show -destination-path jfs_as1:/cg/jfsAA
```

```
Source Path: jfs_as2:/cg/jfsAA
Destination Path: jfs_as1:/cg/jfsAA
Relationship Type: XDP
Relationship Group Type: consistencygroup
SnapMirror Policy Type: automated-failover-duplex
SnapMirror Policy: AutomatedFailOverDuplex
Tries Limit: -
Mirror State: Snapmirrored
Relationship Status: InSync
```

Copyright-Informationen

Copyright © 2026 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtsinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFTE SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGEND EINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

ERLÄUTERUNG ZU „RESTRICTED RIGHTS“: Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

Markeninformationen

NETAPP, das NETAPP Logo und die unter <http://www.netapp.com/TM> aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.