



SnapMirror 활성 동기화

Enterprise applications

NetApp
February 10, 2026

목차

SnapMirror 활성 동기화	1
개요	1
동기식 복제	1
스토리지 하드웨어	1
ONTAP 중재자	1
ONTAP 중재자	1
SnapMirror 활성 동기화 기본 사이트입니다	3
네트워크 토폴로지	4
균일한 액세스	4
비균일 액세스	9
Oracle 구성	11
개요	11
Oracle 단일 인스턴스	11
Oracle RAC 확장	13
RAC 타이브레이커	21
실패 시나리오	22
개요	22
아키텍처 예	24
RAC 상호 연결 오류입니다	25
SnapMirror 통신 실패	26
전체 네트워크 상호 연결 실패	27
사이트 장애	28
중재자 고장	30
서비스 복원	31
수동 페일오버	31

SnapMirror 활성 동기화

개요

SnapMirror Active Sync를 사용하면 서로 다른 두 스토리지 클러스터에서 LUN을 사용할 수 있는 초고가용성 Oracle 데이터베이스 환경을 구축할 수 있습니다.

SnapMirror 활성 동기화에서는 데이터의 "기본" 및 "보조" 복사본이 없습니다. 각 클러스터는 데이터의 로컬 복사본에서 읽기 IO를 제공할 수 있으며, 각 클러스터가 파트너에게 쓰기를 복제합니다. 그 결과는 대칭적인 IO 동작입니다.

이를 통해 Oracle RAC를 두 사이트에서 운영 인스턴스가 있는 확장 클러스터로 실행할 수 있습니다. 또는 RPO=0 Active-Passive 데이터베이스 클러스터를 구축하여 사이트 운영 중단 시 단일 인스턴스 데이터베이스를 사이트 간에 이동할 수 있으며, 이 프로세스는 Pacemaker 또는 VMware HA와 같은 제품을 통해 자동화할 수 있습니다. 이러한 모든 옵션의 기반은 SnapMirror Active Sync로 관리되는 동기식 복제입니다.

동기식 복제

정상 작동 시 SnapMirror Active Sync는 한 가지 경우를 제외하고 항상 RPO=0 동기식 복제본을 제공합니다. 데이터를 복제할 수 없는 경우 ONTAP는 데이터를 복제해야 하는 요구 사항을 해제하고 다른 사이트의 LUN이 오프라인 상태가 되는 동안 한 사이트에서 입출력 서비스를 재개합니다.

스토리지 하드웨어

다른 스토리지 재해 복구 솔루션과 달리 SnapMirror Active Sync는 비대칭적 플랫폼 유연성을 제공합니다. 각 사이트의 하드웨어는 동일할 필요가 없습니다. 이 기능을 사용하면 SnapMirror 액티브 동기화를 지원하는 데 사용되는 하드웨어를 적절한 크기로 조정할 수 있습니다. 전체 운영 워크로드를 지원해야 하는 경우 원격 스토리지 시스템이 기본 사이트와 동일할 수 있지만 재해로 인해 I/O가 감소할 경우 원격 사이트의 소규모 시스템보다 비용 효율적입니다.

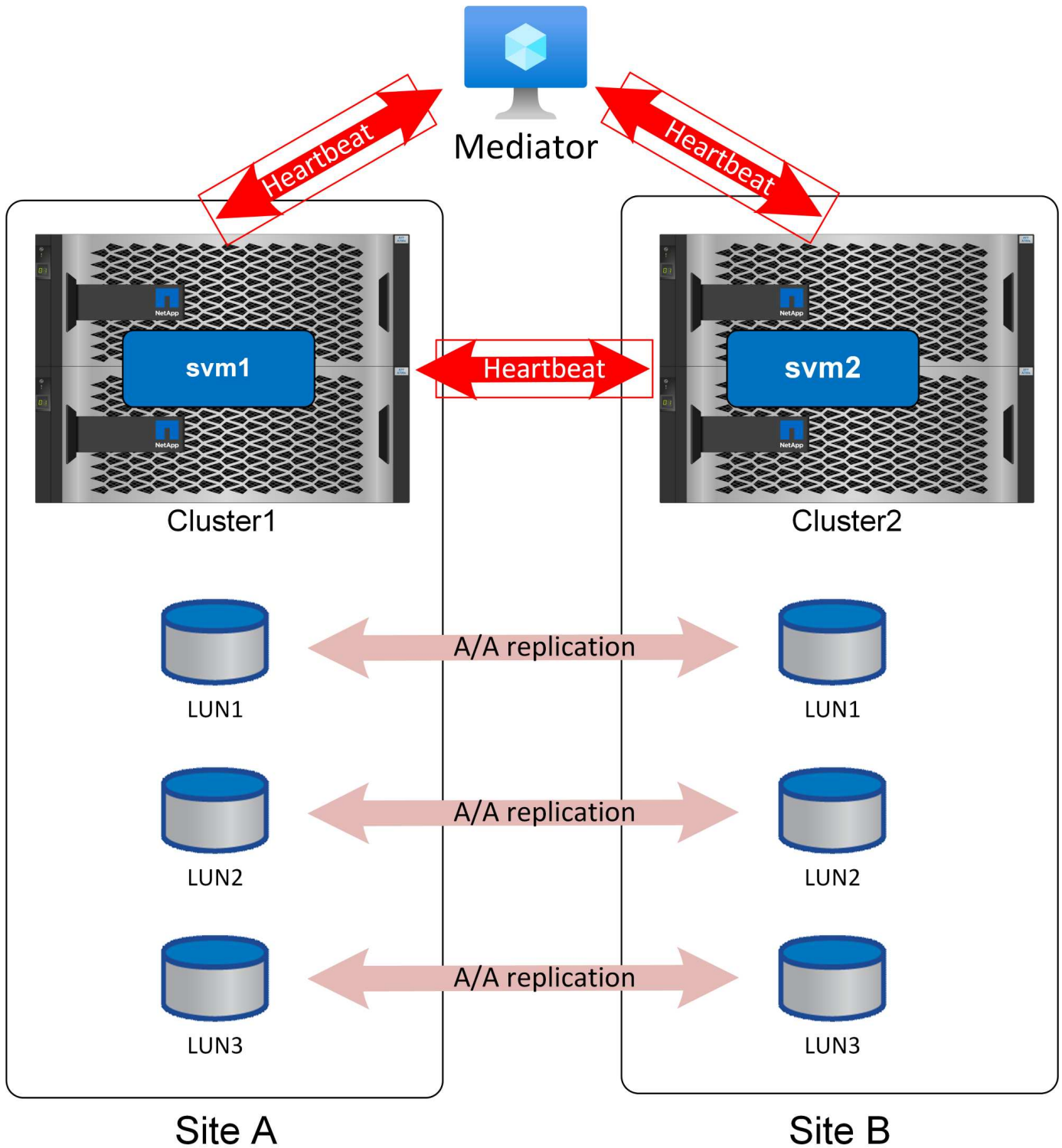
ONTAP 중재자

ONTAP 중재자는 NetApp 지원에서 다운로드되는 소프트웨어 응용 프로그램이며 일반적으로 작은 가상 컴퓨터에 구축됩니다. ONTAP mediator는 SnapMirror active sync와 함께 사용할 때 Tiebreaker가 아닙니다. SnapMirror 활성 동기화 복제에 참여하는 두 클러스터에 대한 대체 통신 채널입니다. 자동화된 운영은 직접 연결과 중재자를 통해 파트너로부터 받은 응답에 따라 ONTAP에 의해 주도됩니다.

ONTAP 중재자

장애 조치를 안전하게 자동화하려면 중재자가 필요합니다. 이상적으로는 독립적인 3차 사이트에 배치되지만 복제에 참여하는 클러스터 중 하나와 같은 위치에 배치하면 대부분의 요구 사항에 대해 여전히 기능을 수행할 수 있습니다.

중재자는 실제로 동점을 깨는 역할을 하지 않지만, 실질적으로는 그런 기능을 제공합니다. 중재자는 클러스터 노드의 상태를 결정하는 데 도움을 주고 사이트 장애 발생 시 자동 전환 프로세스를 지원합니다. 중재자는 어떠한 상황에서도 데이터를 전송하지 않습니다.



자동 장애 조치의 가장 큰 과제는 브레인 분할 문제이며, 두 사이트가 서로 연결이 끊어지면 이 문제가 발생합니다. 어떻게 해야 하나요? 서로 다른 두 사이트가 자신을 데이터의 정상적인 복제본으로 지정하도록 하고 싶지 않지만, 단일 사이트가 상대 사이트의 실제 손실과 반대쪽 사이트와 통신할 수 없는 간의 차이를 어떻게 알 수 있습니까?

중재자가 사진을 입력하는 위치입니다. 세 번째 사이트에 배치되고 각 사이트에 해당 사이트에 대한 별도의 네트워크 연결이 있는 경우 각 사이트에 대한 추가 경로를 통해 다른 사이트의 상태를 확인할 수 있습니다. 위의 그림을 다시 보고 다음 시나리오를 고려하십시오.

- 중재자가 한 사이트 또는 두 사이트에서 작동하지 않거나 연결할 수 없는 경우 어떻게 됩니까?

- 두 클러스터는 복제 서비스에 사용되는 동일한 링크를 통해 서로 계속 통신할 수 있습니다.
- RPO=0 보호로 데이터를 계속 제공합니다
- 사이트 A에 장애가 발생하면 어떻게 됩니까?
 - 사이트 B는 두 통신 채널이 모두 다운되는 것을 볼 수 있습니다.
 - 사이트 B가 데이터 서비스를 인수하지만 RPO = 0 미러링이 없습니다
- 사이트 B에 장애가 발생하면 어떻게 됩니까?
 - 사이트 A는 두 통신 채널이 모두 다운되는 것을 볼 수 있습니다.
 - 사이트 A가 데이터 서비스를 인수하지만 RPO = 0 미러링이 없음

고려해야 할 다른 시나리오가 있습니다: 데이터 복제 링크의 손실. 사이트 간 복제 링크가 손실되면 RPO=0 미러링이 불가능할 것입니다. 그러면 어떻게 됩니까?

이는 선호 사이트 상태에 의해 제어됩니다. SM-AS 관계에서 사이트 중 하나가 다른 사이트에 대한 보조 사이트입니다. 이 작업은 일반 작업에는 영향을 주지 않으며 모든 데이터 액세스는 대칭이지만 복제가 중단되면 작업을 재개하려면 연결을 끊어야 합니다. 그 결과 기본 사이트가 미러링 없이 운영을 계속하고 복제 통신이 복구될 때까지 보조 사이트가 입출력 처리를 중지합니다.

SnapMirror 활성 동기화 기본 사이트입니다

SnapMirror 액티브 동기화 동작은 대칭이며 한 가지 중요한 예외 기본 사이트 구성이 있습니다.

SnapMirror 활성 동기화는 한 사이트를 "소스"로 간주하고 다른 사이트는 "대상"으로 간주합니다. 이는 단방향 복제 관계를 의미하지만 입출력 동작에는 적용되지 않습니다. 복제는 양방향이고 대칭이며 입출력 응답 시간은 미러의 양쪽에서 동일합니다.

`source` 지정은 기본 사이트를 제어합니다. 복제 링크가 손실되면 소스 복제본의 LUN 경로는 계속 데이터를 제공하고 대상 복제본의 LUN 경로는 복제가 다시 설정되고 SnapMirror가 동기식 상태로 다시 전환될 때까지 사용할 수 없게 됩니다. 그러면 경로가 데이터 제공을 재개합니다.

소스/대상 구성은 SystemManager를 통해 볼 수 있습니다.

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

또는 CLI에서:

```
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
```

핵심은 소스는 클러스터 1의 SVM입니다. 위에서 언급한 바와 같이 "원본" 및 "대상"이라는 용어는 복제된 데이터의 흐름을 설명하지 않습니다. 두 사이트 모두 쓰기를 처리하여 반대쪽 사이트로 복제할 수 있습니다. 실제로 두 클러스터 모두 소스와 타겟입니다. 하나의 클러스터를 소스로 지정하면 복제 링크가 손실된 경우 읽기-쓰기 스토리지 시스템으로 존속하는 클러스터를 제어할 수 있습니다.

네트워크 토폴로지

균일한 액세스

균일한 액세스 네트워킹은 호스트가 두 사이트(또는 동일한 사이트 내의 장애 도메인)의 경로를 액세스할 수 있음을 의미합니다.

SM-AS의 중요한 기능은 호스트의 위치를 알 수 있도록 스토리지 시스템을 구성하는 기능입니다. 특정 호스트에 LUN을 매핑할 때 해당 LUN이 지정된 스토리지 시스템에 근접한지 여부를 지정할 수 있습니다.

근접 설정

근접성은 특정 호스트 WWN 또는 iSCSI 이니시에이터 ID가 로컬 호스트에 속함을 나타내는 클러스터별 구성을 의미합니다. LUN 액세스를 구성하는 두 번째 선택적 단계입니다.

첫 번째 단계는 일반적인 igroup 구성입니다. 각 LUN은 해당 LUN에 액세스해야 하는 호스트의 WWN/iSCSI ID가 포함된 igroup에 매핑되어야 합니다. LUN에 대한 액세스 권한이 있는 호스트를 제어합니다.

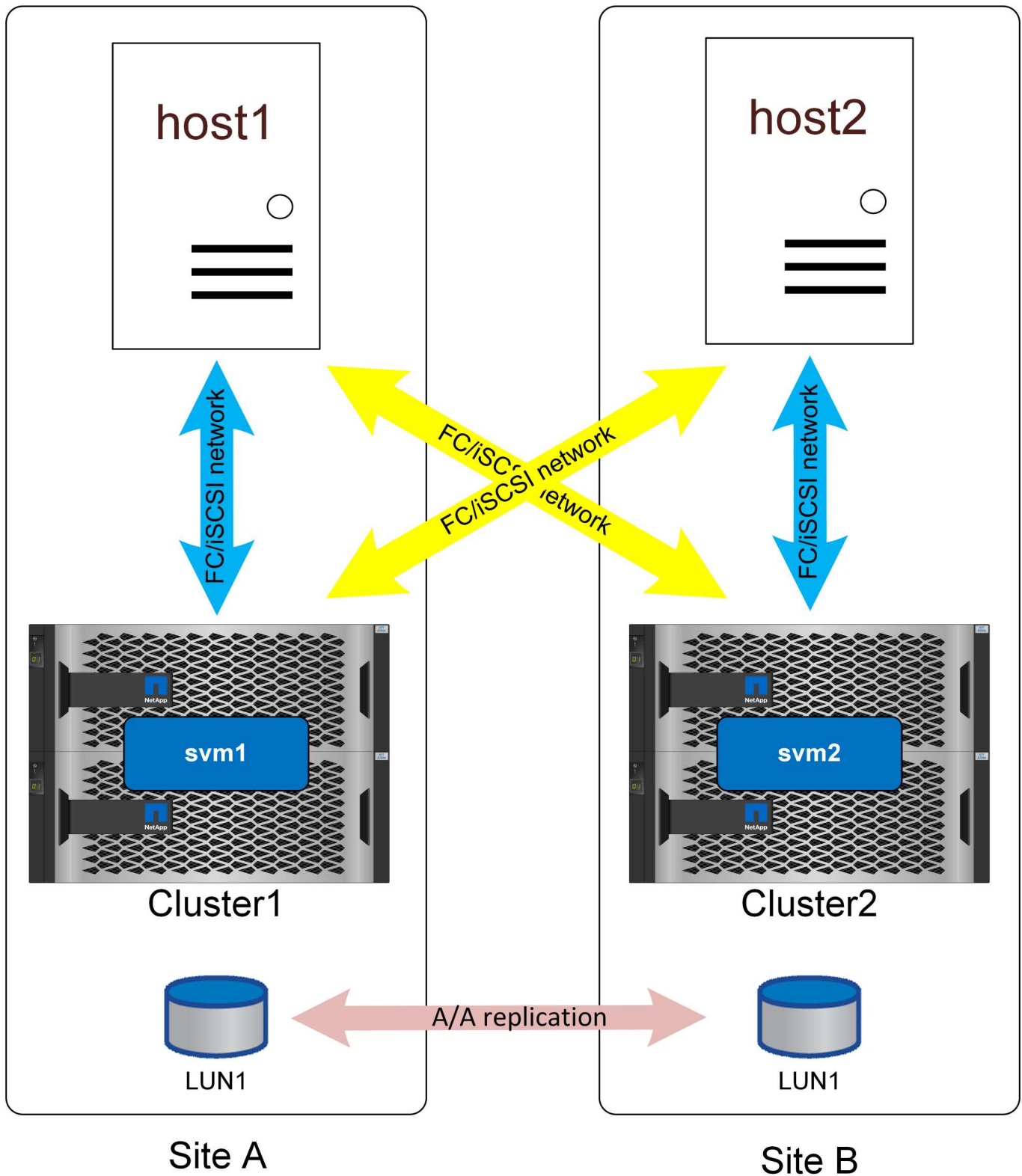
두 번째 단계는 호스트 근접성을 구성하는 것입니다. 액세스를 제어하지 않고 `_priority_`를 제어합니다.

예를 들어 사이트 A의 호스트가 SnapMirror 활성 동기화로 보호되는 LUN을 액세스하도록 구성할 수 있고 SAN이 사이트 간에 확장되므로 사이트 A의 스토리지 또는 사이트 B의 스토리지를 사용하여 해당 LUN에 대한 경로를 사용할 수 있습니다.

근접 설정이 없으면 두 스토리지 시스템 모두 활성/최적화된 경로를 홍보하기 때문에 해당 호스트는 두 스토리지 시스템을 동일하게 사용합니다. 사이트 간 SAN 대기 시간 및/또는 대역폭이 제한된 경우 이를 원하지 않을 수 있으며, 정상적인 작업 중에 각 호스트가 로컬 스토리지 시스템에 대한 경로를 우선적으로 사용하도록 할 수 있습니다. 호스트 WWN/iSCSI ID를 로컬 클러스터에 근위부 호스트로 추가하여 구성합니다. 이 작업은 CLI 또는 SystemManager에서 수행할 수 있습니다.

AFF

AFF 시스템의 경우 호스트 근접성이 구성된 경우 경로가 아래와 같이 표시됩니다.



Active/Optimized Path

Active Path

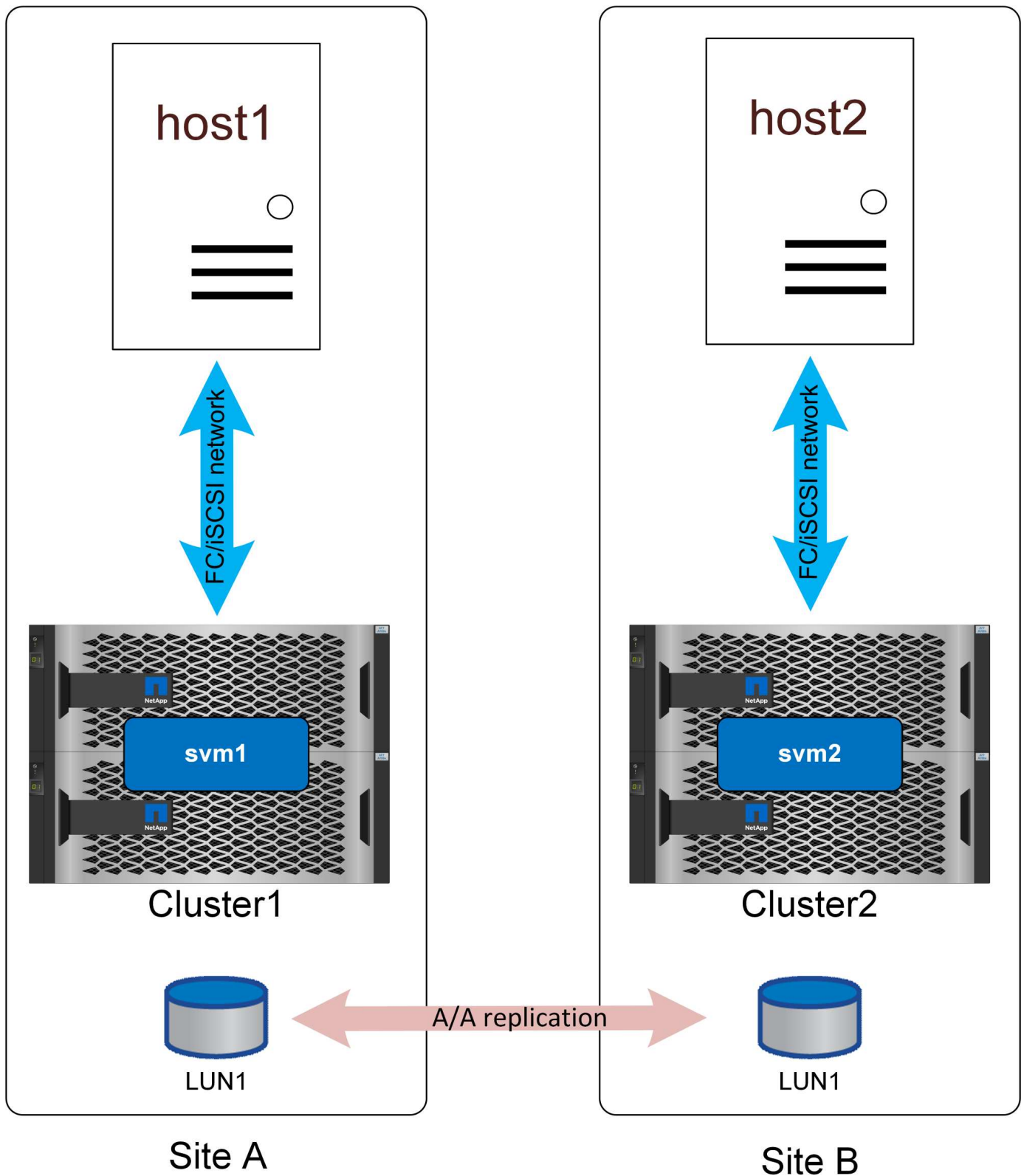
정상 작동 시 모든 입출력은 로컬 입출력입니다. 읽기와 쓰기는 로컬 스토리지 시스템에서 서비스됩니다. 물론 쓰기 입출력도 인식되기 전에 로컬 컨트롤러에 의해 원격 시스템으로 복제되어야 하지만 모든 읽기 입출력은 로컬로 처리되며 사이트 간 SAN 링크를 통과하여 추가적인 지연 시간이 발생하지 않습니다.

최적화되지 않은 경로는 액티브/최적화 경로가 모두 손실되는 경우에만 사용됩니다. 예를 들어, 사이트 A의 전체 스토리지에 전원이 공급되지 않으면 사이트 A의 호스트가 사이트 B의 어레이에 대한 경로에 액세스할 수 있으므로 지연 시간이 더 길더라도 작동 상태를 유지할 수 있습니다.

로컬 클러스터를 통한 중복 경로가 있습니다. 이러한 경로는 단순성을 위해 이러한 다이어그램에 표시되지 않습니다. ONTAP 스토리지 시스템은 HA 자체이므로 컨트롤러 장애가 사이트 장애로 이어질 수 없습니다. 영향을 받는 사이트에서 사용되는 로컬 경로가 변경될 뿐입니다.

ASA

NetApp ASA 시스템은 클러스터의 모든 경로에 대해 액티브-액티브 다중 경로를 제공합니다. 이는 SM-AS 구성에도 적용됩니다.



Active/Optimized Path

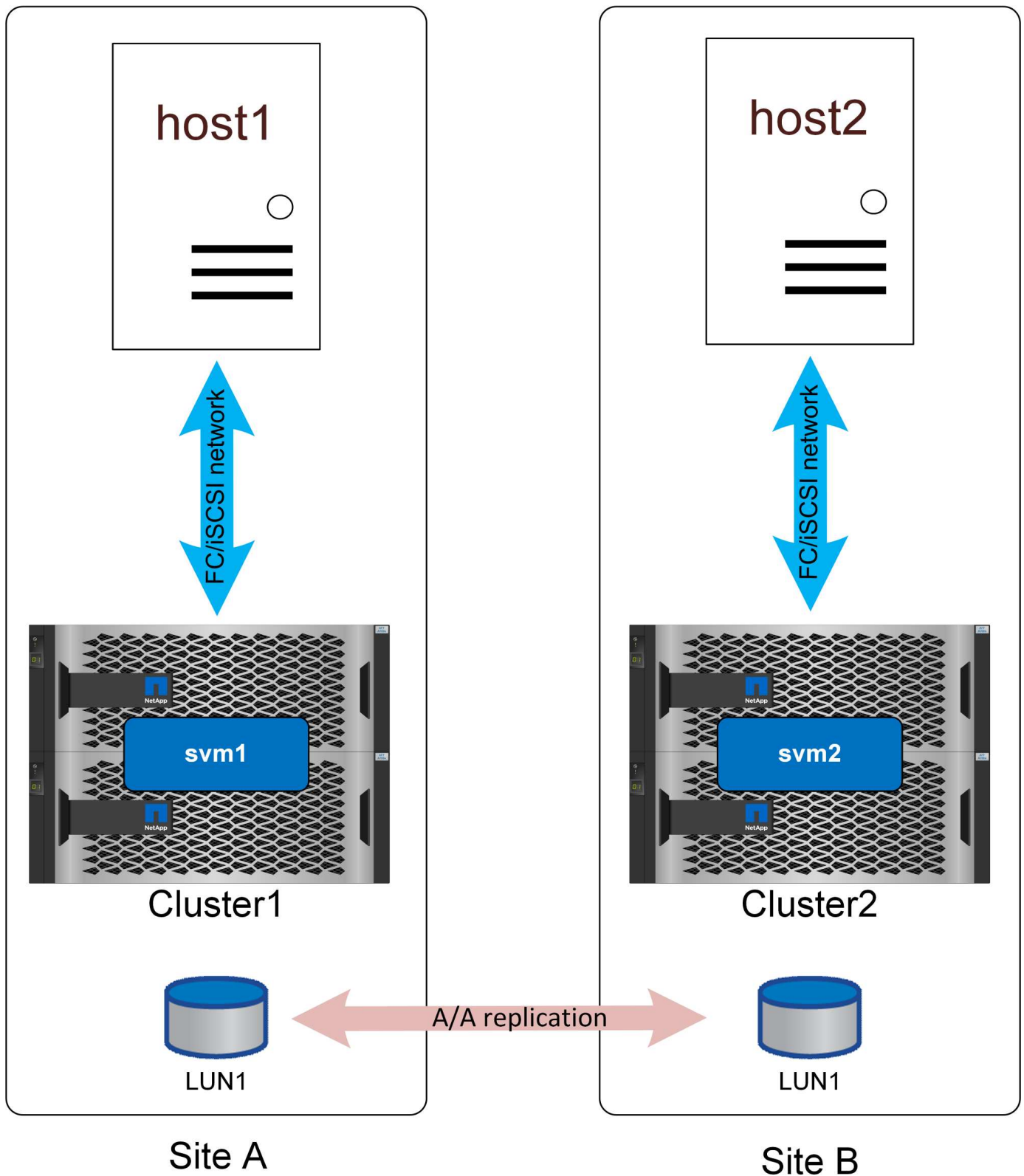
비균일 액세스를 사용하는 ASA 구성은 AFF와 거의 동일합니다. 균일한 액세스의 경우 입출력이 WAN을 통과하게 됩니다. 이것은 바람직하지 않을 수도 있습니다.

두 사이트가 파이버 연결을 통해 100m 떨어져 있는 경우 WAN을 통해 추가 지연 시간을 감지할 수 없지만 사이트가 멀리 떨어져 있으면 두 사이트에서 읽기 성능이 저하됩니다. 이와 반대로 AFF에서는 사용 가능한 로컬 경로가 없는 경우에만 WAN 교차 경로를 사용하고 모든 입출력이 로컬 입출력이기 때문에 일상적인 성능이 더 나을 것입니다. 비균일 액세스 네트워크를 사용하는 ASA는 사이트 간 지연 시간 액세스 패널티 없이 ASA의 비용 및 기능 이점을 얻을 수 있는 옵션이 될 것입니다.

지연 시간이 짧은 구성의 SM-AS를 사용하는 ASA는 두 가지 흥미로운 이점을 제공합니다. 첫째로, IO는 2배 더 많은 경로를 사용하여 2배 더 많은 컨트롤러가 처리할 수 있기 때문에 모든 단일 호스트의 성능을 두 배 * 제공합니다. 둘째, 단일 사이트 환경에서 호스트 액세스를 중단하지 않고 전체 스토리지 시스템을 손실할 수 있기 때문에 최고의 가용성을 제공합니다.

비균일 액세스

비균일 액세스 네트워킹은 각 호스트가 로컬 스토리지 시스템의 포트에만 액세스할 수 있음을 의미합니다. SAN은 사이트(또는 동일한 사이트 내의 장애 도메인)에 걸쳐 확장되지 않습니다.



Active/Optimized Path

이 접근 방식의 주요 이점은 SAN의 단순성입니다. 네트워크를 통해 SAN을 확장할 필요가 없어졌습니다. 일부 고객은 사이트 간에 지연 시간이 충분히 짧거나 사이트 간 네트워크를 통해 FC SAN 트래픽을 터널링할 수 있는 인프라가

부족합니다.

비균일 액세스의 단점은 복제 링크 손실을 비롯한 특정 장애 시나리오로 인해 일부 호스트가 스토리지에 액세스할 수 없게 된다는 것입니다. 클러스터 이외의 데이터베이스와 같이 단일 인스턴스로 실행되는 애플리케이션은 로컬 스토리지 연결이 끊긴 경우 해당 마운트의 단일 호스트에서만 실행 중이므로 실패합니다. 데이터는 여전히 보호되지만 데이터베이스 서버는 더 이상 액세스할 수 없습니다. 원격 사이트에서 자동 프로세스를 통해 재시작해야 합니다. 예를 들어 VMware HA는 한 서버에서 모든 경로 다운 상황을 감지하고 경로를 사용할 수 있는 다른 서버에서 VM을 다시 시작할 수 있습니다.

반면, Oracle RAC와 같은 클러스터된 애플리케이션은 두 개의 사이트에서 동시에 사용할 수 있는 서비스를 제공할 수 있습니다. 사이트를 잃었다고 해서 애플리케이션 서비스 전체가 손실되는 것은 아닙니다. 인스턴스는 계속 사용할 수 있으며 정상적인 사이트에서 실행됩니다.

대부분의 경우, 사이트 간 링크를 통해 스토리지에 액세스하는 애플리케이션의 지연 시간 오버헤드가 허용할 수 없는 경우가 많습니다. 즉, 사이트의 스토리지 손실로 인해 장애가 발생한 사이트의 서비스를 종료해야 하기 때문에 일관된 네트워킹의 가용성 향상은 최소화됩니다.



로컬 클러스터를 통한 중복 경로가 있습니다. 이러한 경로는 단순성을 위해 이러한 다이어그램에 표시되지 않습니다. ONTAP 스토리지 시스템은 HA 자체이므로 컨트롤러 장애가 사이트 장애로 이어질 수 없습니다. 영향을 받는 사이트에서 사용되는 로컬 경로가 변경될 뿐입니다.

Oracle 구성

개요

SnapMirror 활성 동기화를 사용한다고 해서 데이터베이스 운영에 대한 모범 사례가 반드시 추가되거나 변경되는 것은 아닙니다.

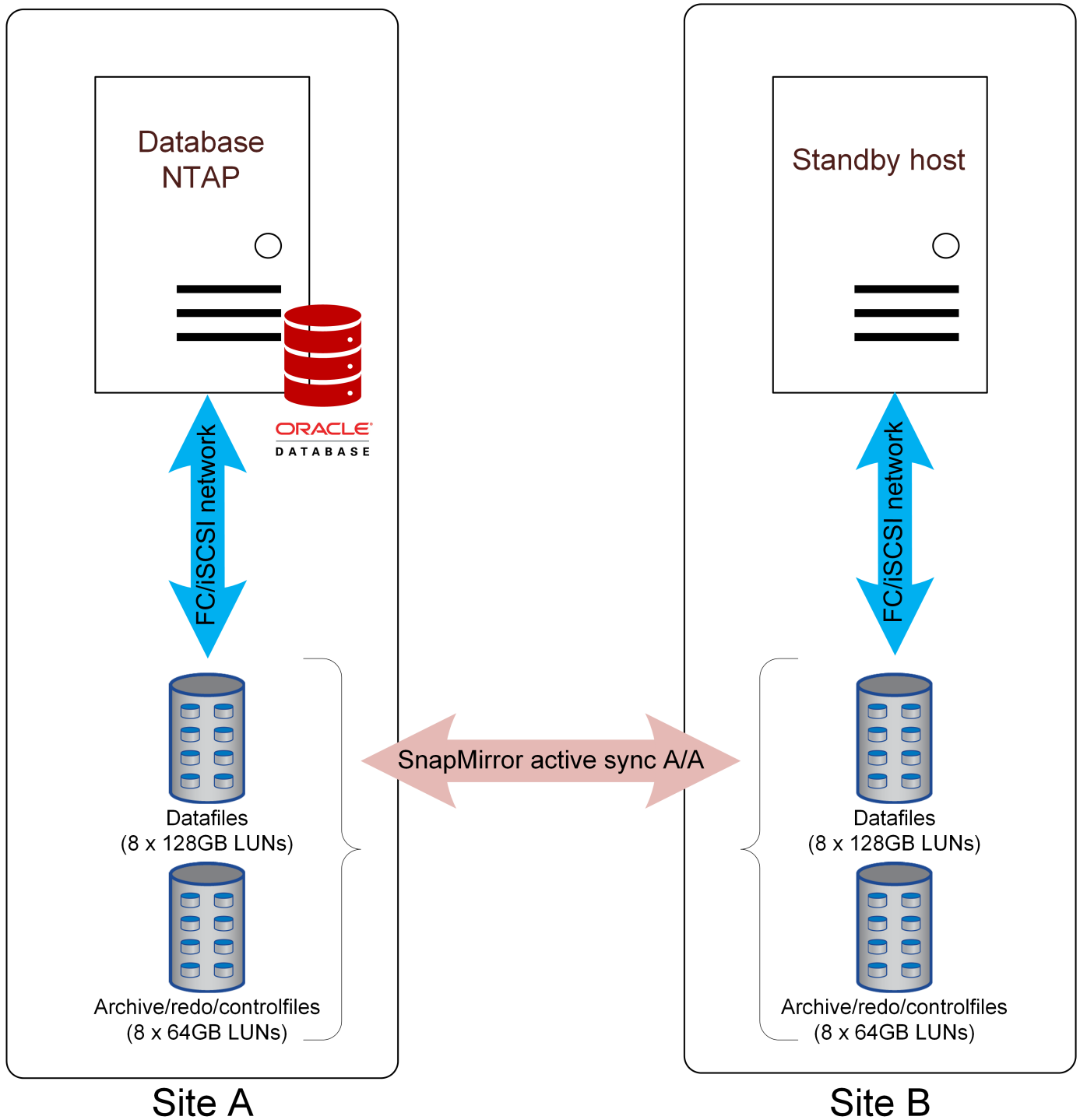
최상의 아키텍처는 비즈니스 요구 사항에 따라 달라집니다. 예를 들어 RPO=0 데이터 손실에 대한 보호를 목표로 하지만 RTO가 완화되는 경우 Oracle 단일 인스턴스 데이터베이스를 사용하고 SM-AS로 LUN을 복제하는 것만으로도 충분할 뿐만 아니라 Oracle 라이선스 스탠더딩 방식에서 비용이 적게 들 수 있습니다. 원격 사이트에 장애가 발생해도 작업이 중단되지 않으며, 운영 사이트가 손실되면 정상적인 사이트에서 LUN을 사용할 수 있는 상태로 온라인 상태로 유지됩니다.

RTO가 보다 엄격하다면 페이스 메이커나 Ansible과 같은 클러스터웨어나 스크립트를 통한 기본 액티브-패시브 자동화를 통해 페일오버 시간을 향상할 수 있습니다. 예를 들어 운영 사이트에서 VM 장애를 감지하고 원격 사이트에서 VM을 활성화하도록 VMware HA를 구성할 수 있습니다.

마지막으로, 매우 신속한 페일오버를 위해 Oracle RAC를 여러 사이트에 구축할 수 있습니다. RTO는 기본적으로 0입니다. 데이터베이스가 온라인 상태이고 두 사이트에서 항상 사용할 수 있기 때문입니다.

Oracle 단일 인스턴스

아래에 설명된 예는 SnapMirror 활성 동기화 복제를 사용하여 Oracle 단일 인스턴스 데이터베이스를 구축하는 여러 가지 옵션 중 일부를 보여 줍니다.



사전 구성된 OS로 페일오버

SnapMirror 액티브 동기화는 재해 복구 사이트에서 데이터의 동기식 복제본을 제공하지만, 해당 데이터를 사용하려면 운영 체제와 관련 애플리케이션이 필요합니다. 기본 자동화를 통해 전체 환경의 장애 조치 시간을 크게 개선할 수 있습니다. Pacemaker와 같은 Clusterware 제품은 사이트 전체에 클러스터를 생성하는 데 자주 사용되며, 대부분의 경우 간단한 스크립트로 장애 조치 프로세스를 구동할 수 있습니다.

운영 노드가 손실되면 클러스터웨어(또는 스크립트)가 대체 사이트에서 데이터베이스를 온라인으로 전환합니다. 한 가지 옵션은 데이터베이스를 구성하는 SAN 리소스에 대해 미리 구성된 대기 서버를 만드는 것입니다. 운영 사이트에 장애가 발생하면 클러스터웨어 또는 스크립트된 대체 시스템이 다음과 유사한 일련의 작업을 수행합니다.

1. 운영 사이트의 장애를 감지합니다
2. FC 또는 iSCSI LUN 검색 수행
3. 파일 시스템 마운트 및/또는 ASM 디스크 그룹 마운트
4. 데이터베이스를 시작하는 중입니다

이 방법의 주요 요구 사항은 원격 사이트에서 실행 중인 OS입니다. Oracle 바이너리로 사전 구성되어야 합니다. 즉, Oracle 패치 적용과 같은 작업이 운영 및 대기 사이트에서 수행되어야 합니다. 또는 재해가 선언된 경우 Oracle 바이너리를 원격 사이트로 미러링하고 마운트할 수 있습니다.

실제 활성화 절차는 간단합니다. LUN 검색과 같은 명령은 FC 포트당 몇 개의 명령만 사용하면 됩니다. 파일 시스템 마운팅은 명령에 불과하며 mount, 단일 명령으로 CLI에서 데이터베이스와 ASM을 모두 시작하고 중지할 수 있습니다.

가상화된 OS로 페일오버

데이터베이스 환경의 페일오버는 운영 체제 자체를 포함하도록 확장할 수 있습니다. 이론적으로 이 페일오버는 부팅 LUN에서 수행할 수 있지만 대부분의 경우 가상화된 OS에서 수행됩니다. 절차는 다음 단계와 유사합니다.

1. 운영 사이트의 장애를 감지합니다
2. 데이터베이스 서버 가상 머신을 호스팅하는 데이터 저장소를 마운트합니다
3. 가상 머신 시작
4. 데이터베이스를 수동으로 시작하거나 가상 시스템이 데이터베이스를 자동으로 시작하도록 구성합니다.

예를 들어, ESX 클러스터가 사이트에 걸쳐 있을 수 있습니다. 재해 발생 시 전환 후 재해 복구 사이트에서 가상 시스템을 온라인으로 전환할 수 있습니다.

스토리지 장애 보호

위의 다이어그램은 "비균일 액세스" SAN이 사이트 간에 확장되지 않는 의 사용을 보여 줍니다. 이 방법은 구성이 더 간단할 수 있으며 경우에 따라 현재 SAN 기능을 고려할 때 유일한 옵션이 될 수도 있지만 이는 또한 운영 스토리지 시스템의 장애로 인해 애플리케이션이 페일오버될 때까지 데이터베이스가 중단된다는 것을 의미합니다.

복원력을 높이기 위해 이 솔루션을 과 함께 구축할 수 "균일한 액세스" 있습니다. 이렇게 하면 반대쪽 사이트에서 알리는 경로를 사용하여 응용 프로그램이 계속 작동할 수 있습니다.

Oracle RAC 확장

많은 고객이 사이트 간에 Oracle RAC 클러스터를 확장하여 완벽한 Active-Active 구성을 실현함으로써 RTO를 최적화합니다. Oracle RAC의 쿼럼 관리를 포함해야 하기 때문에 전체 설계가 더 복잡해집니다.

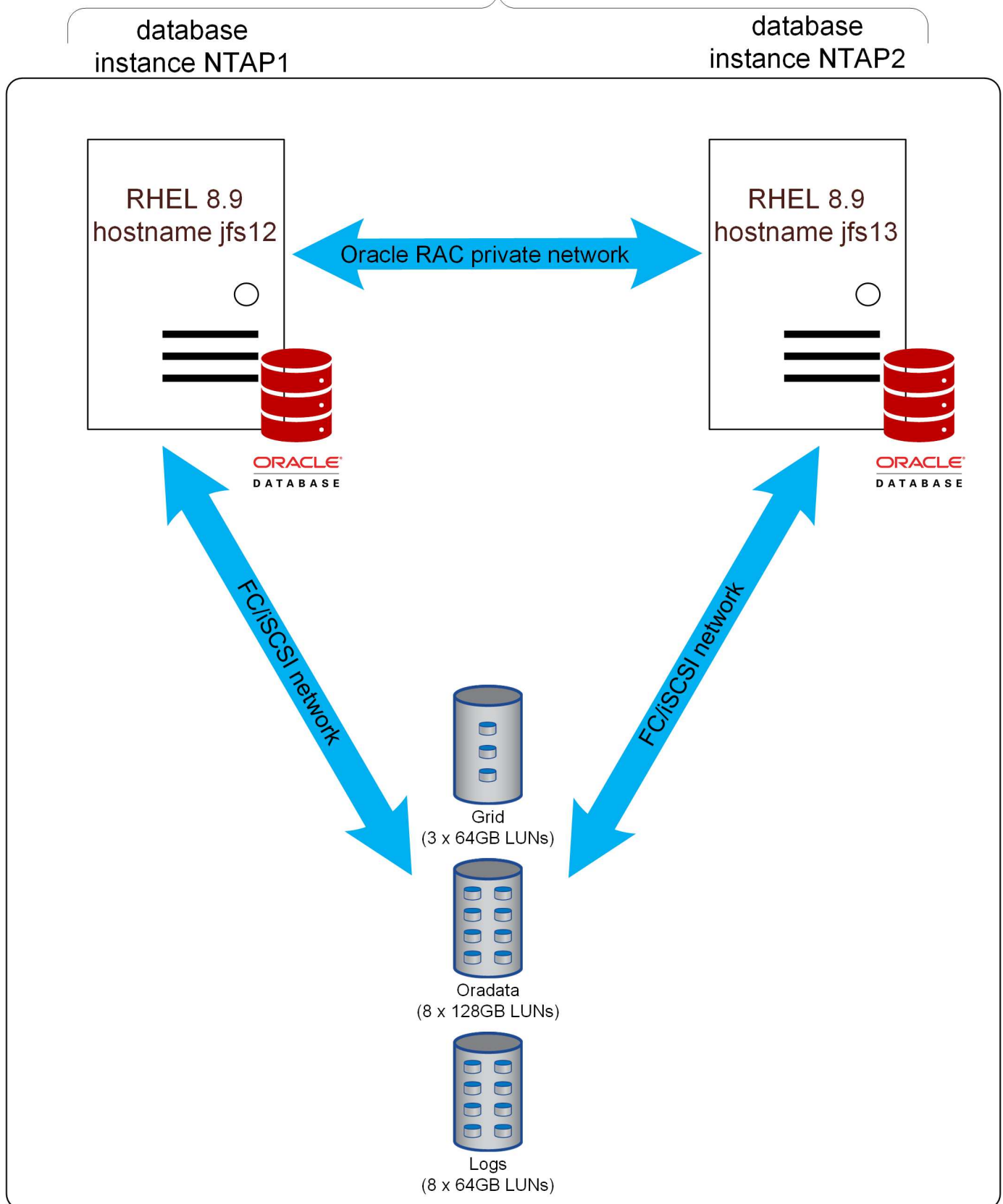
기존의 확장된 RAC 클러스터에서는 ASM 미러링에 의존하여 데이터 보호를 제공했습니다. 이 접근 방식은 효과가 있지만 많은 수동 구성 단계가 필요하며 네트워크 인프라에 오버헤드를 발생시킵니다. 반면, SnapMirror Active Sync가 데이터 복제에 대한 책임을 지도록 하면 솔루션이 크게 간소화됩니다. 동기화, 중단 후 재동기화, 장애 복구, 쿼럼 관리 등의 작업이 더 쉬워지고 SAN을 여러 사이트에 분산할 필요가 없어 SAN 설계 및 관리가 간소화됩니다.

복제

SnapMirror Active Sync의 RAC 기능을 이해하는 데 중요한 것은 미러링된 스토리지에서 호스팅되는 단일 LUN

세트로 스토리지를 보는 것입니다. 예를 들면 다음과 같습니다.

Database NTAP



운영 복사본 또는 미러 복사본이 없습니다. 논리적으로, 각 LUN에는 단일 복제본만 있으며, 이 LUN은 서로 다른 두 스토리지 시스템에 있는 SAN 경로에서 사용할 수 있습니다. 호스트 관점에서 볼 때 스토리지 페일오버가 발생하지 않고 경로 변경이 있습니다. 다양한 장애 이벤트로 인해 다른 경로가 온라인 상태로 유지되는 동안 LUN에 대한 특정 경로가 손실될 수 있습니다. SnapMirror 활성 동기화는 모든 운영 경로에 대해 동일한 데이터를 사용할 수 있도록 보장합니다.

스토리지 구성

이 예제 구성에서는 ASM 디스크가 엔터프라이즈 스토리지의 단일 사이트 RAC 구성과 동일하게 구성됩니다. 스토리지 시스템이 데이터 보호를 제공하기 때문에 ASM 외부 중복성이 사용됩니다.

균일 액세스 대 비정보 액세스

SnapMirror 액티브 동기화에서 Oracle RAC와 관련하여 가장 중요한 고려사항은 균일한 액세스를 사용하는지 비균일한 액세스를 사용하는지 여부입니다.

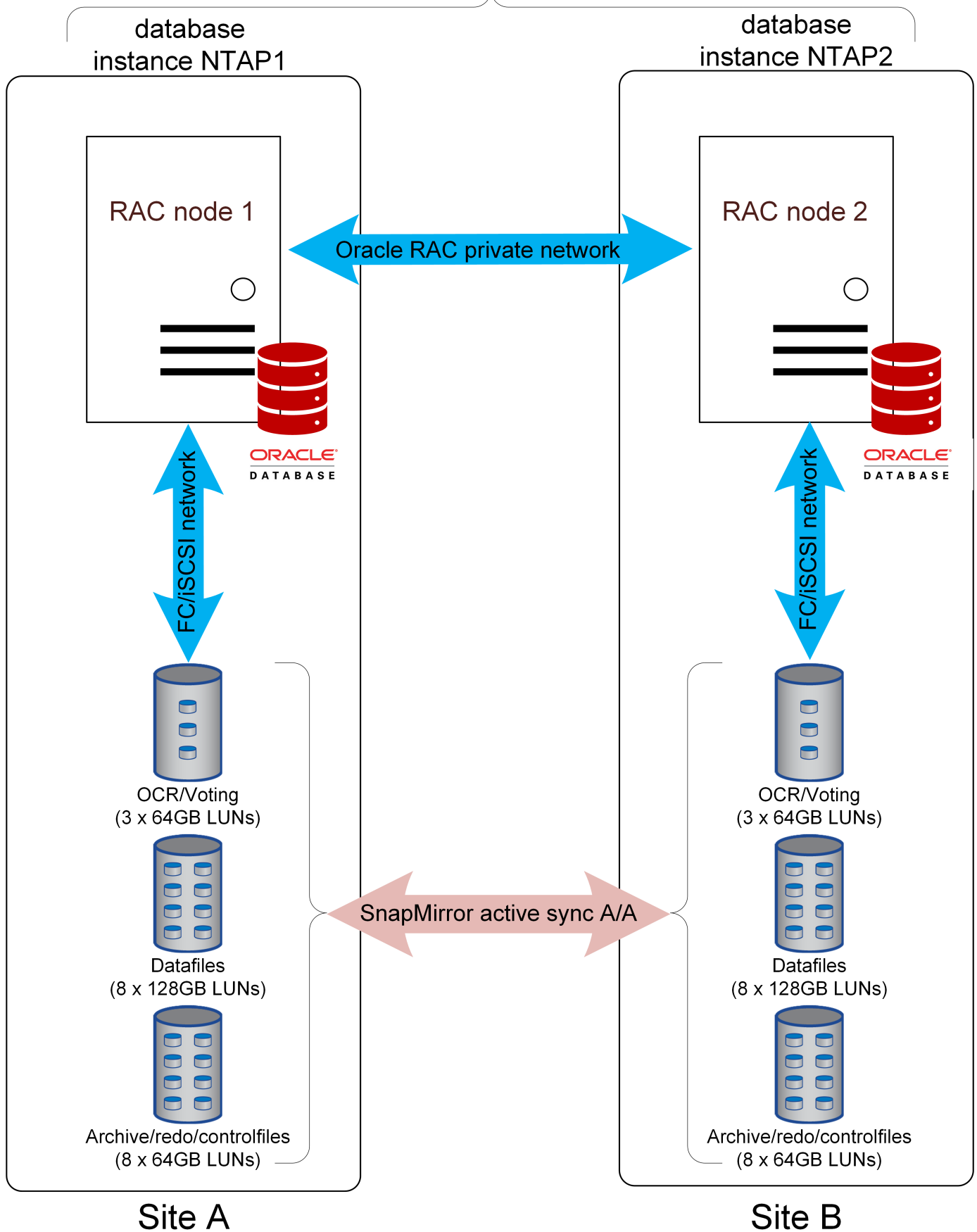
균일한 액세스는 각 호스트가 두 클러스터 모두에서 경로를 볼 수 있음을 의미합니다. 비균일 액세스는 호스트가 로컬 클러스터에 대한 경로만 볼 수 있음을 의미합니다.

두 옵션은 특별히 권장되거나 권장되지 않습니다. 일부 고객은 사이트에 쉽게 연결할 수 있는 다크 파이버를 보유하고 있으며, 다른 고객은 이러한 연결을 보유하고 있지 않거나 SAN 인프라스트럭처가 장거리 ISL을 지원하지 않습니다.

비균일 액세스

비균일 액세스는 SAN 관점에서 구성하는 것이 더 간단합니다.

Database NTAP



기본 단점은 "비균일 액세스"사이트 간 ONTAP 연결이 끊어지거나 스토리지 시스템이 손실되면 한 사이트에서 데이터베이스 인스턴스가 손실된다는 것입니다. 이는 바람직한 방법은 아니지만 좀 더 단순한 SAN 구성을 위한 대가로 허용 가능한 위험일 수 있습니다.

균일한 액세스

단일 액세스를 사용하려면 사이트 간에 SAN을 확장해야 합니다. 주된 이점은 스토리지 시스템의 손실로 인해 데이터베이스 인스턴스가 손실되지는 않는다는 것입니다. 대신, 경로가 현재 사용 중인 다중 경로가 변경됩니다.

여러 가지 방법으로 비균일 액세스를 구성할 수 있습니다.

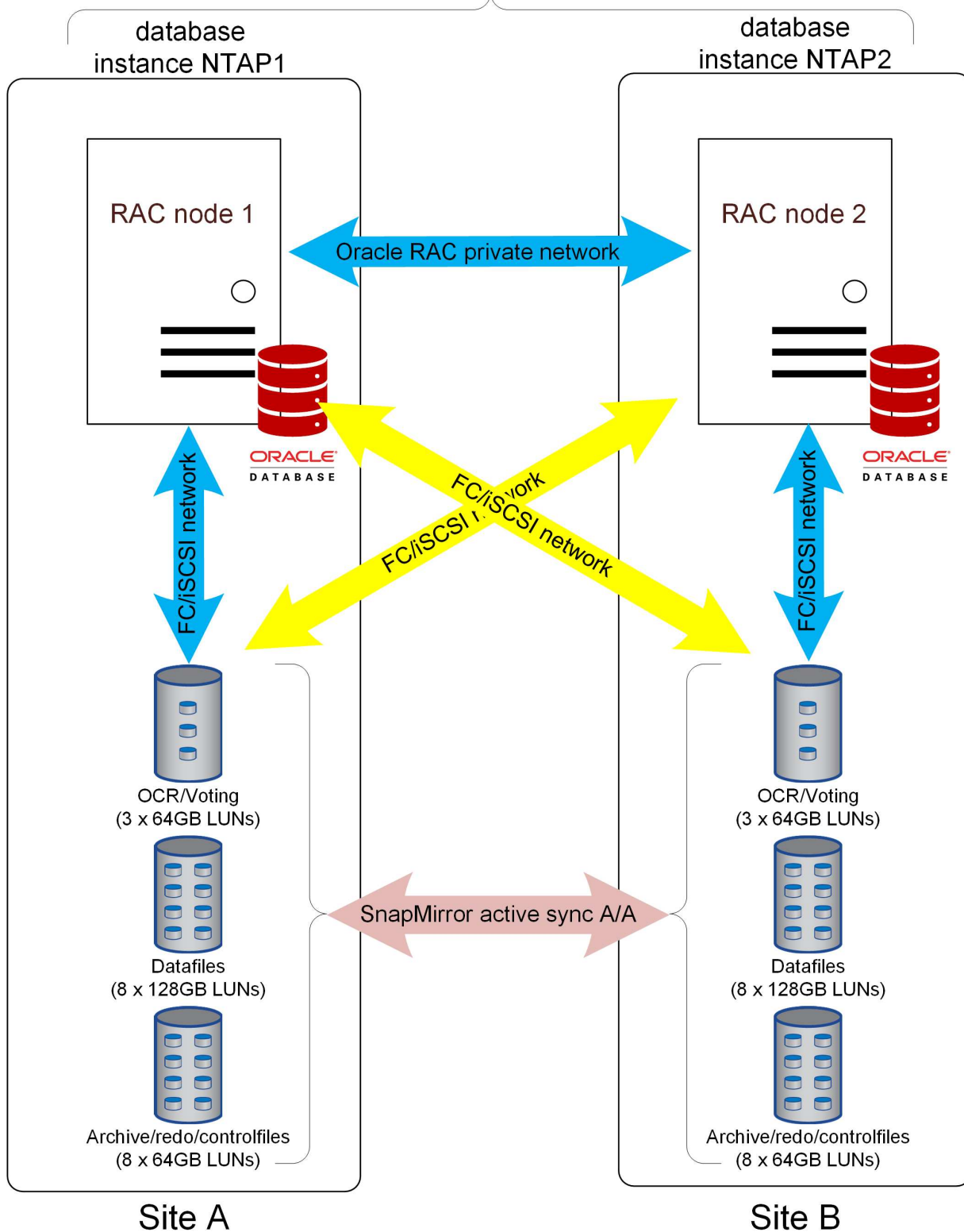


아래 다이어그램에는 단순한 컨트롤러 장애 시 사용되는 활성 경로이지만 최적화되지 않은 경로가 있지만 다이어그램을 단순화하기 위해 이러한 경로가 표시되지 않습니다.

근접 설정이 있는 AFF

사이트 간에 지연 시간이 클 경우 호스트 근접 설정을 사용하여 AFF 시스템을 구성할 수 있습니다. 따라서 각 스토리지 시스템에서 로컬인 호스트와 원격인 호스트를 인식하고 경로 우선 순위를 적절하게 할당할 수 있습니다.

Database NTAP



Active/Optimized Path

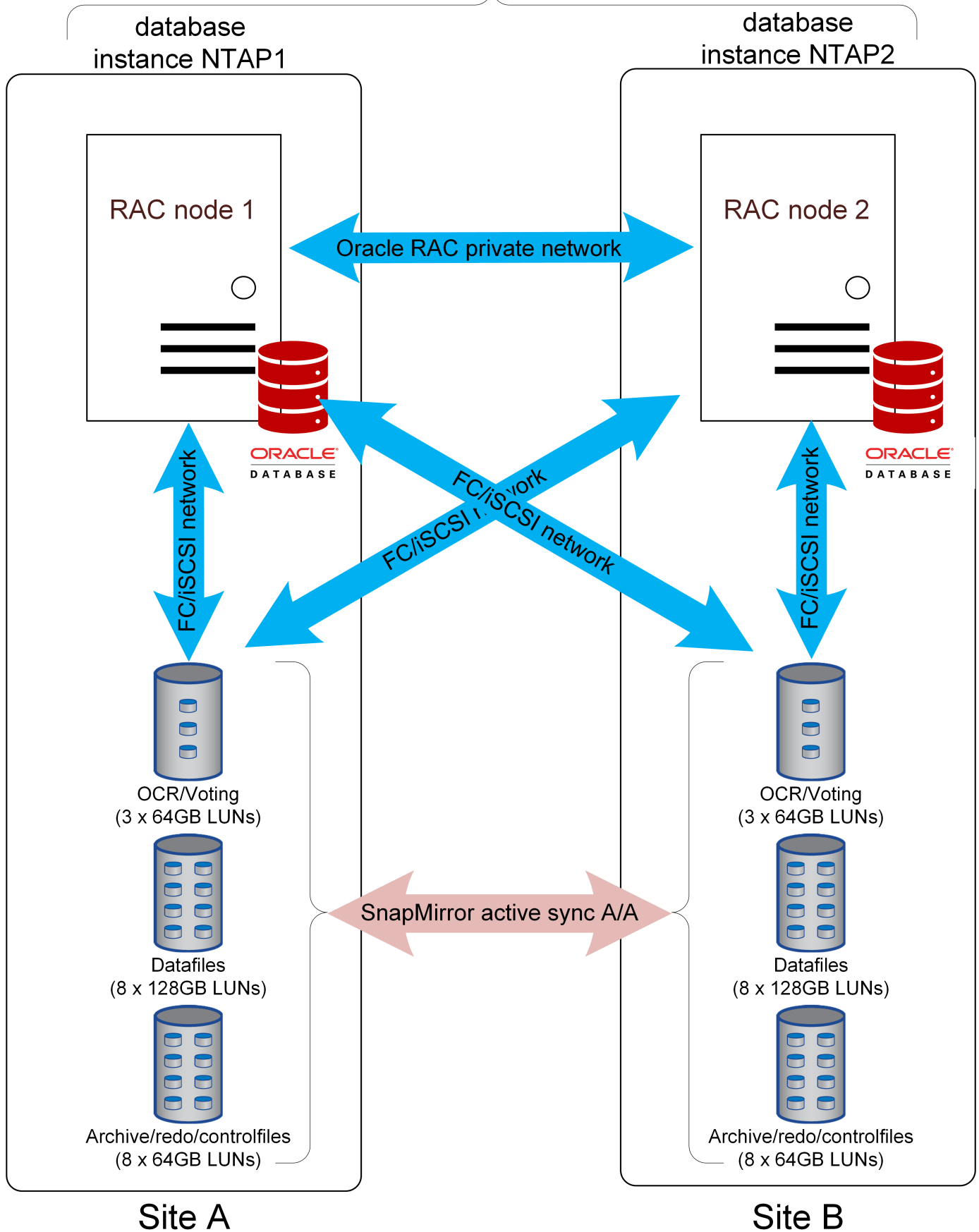
Active Path

정상 작동 시 각 Oracle 인스턴스는 로컬 활성/최적화된 경로를 우선적으로 사용합니다. 결과적으로 블록의 로컬 복제본이 모든 읽기를 서비스합니다. 따라서 지연 시간이 가장 짧아집니다. 마찬가지로 쓰기 입출력도 로컬 컨트롤러로 하향 전송됩니다. 입출력이 확인되기 전에 복제되어야 하므로 사이트 간 네트워크를 가로지르는 데 추가적인 지연 시간이 발생하지만 동기식 복제 솔루션에서는 이를 방지할 수 없습니다.

근접 설정이 없는 **ASA/AFF**

사이트 간에 심각한 지연 시간이 없다면 호스트 근접 설정 없이 AFF 시스템을 구성하거나 ASA를 사용할 수 있습니다.

Database NTAP



각 호스트는 두 스토리지 시스템에서 모든 운영 경로를 사용할 수 있습니다. 따라서 각 호스트가 하나의 클러스터가 아닌 두 클러스터의 성능 잠재력을 활용할 수 있으므로 성능이 크게 향상됩니다.

ASA를 사용하면 두 클러스터에 대한 모든 경로를 활성 및 최적화된 것으로 간주할 뿐만 아니라 파트너 컨트롤러의 경로도 활성 상태로 간주됩니다. 그 결과 항상 전체 클러스터에서 All-Active SAN 경로가 됩니다.



ASA 시스템은 비균일 액세스 구성에서도 사용할 수 있습니다. 사이트 간 경로가 없으므로 ISL을 통한 입출력으로 인해 성능에 영향을 미치지 않습니다.

RAC 타이브레이커

SnapMirror 액티브 동기화를 사용하는 확장된 RAC는 IO와 관련된 대칭 아키텍처이지만 브레인 분할 관리에 연결된 한 가지 예외가 있습니다.

복제 링크가 손실되고 양쪽 사이트에 쿼럼이 없는 경우 어떻게 됩니까? 어떻게 해야 합니까? 이 질문은 Oracle RAC 및 ONTAP 동작 모두에 적용됩니다. 변경 내용을 사이트 간에 복제할 수 없고 작업을 다시 시작하려는 경우 사이트 중 하나가 생존해야 하고 다른 사이트를 사용할 수 없게 됩니다.

은 "[ONTAP 중재자](#)" ONTAP 계층에서 이 요구 사항을 해결합니다. RAC 티브레이킹에는 여러 옵션이 있습니다.

Oracle Tiebreaker를 소개합니다

브레인 분할 Oracle RAC 위험을 관리하기 위한 가장 좋은 방법은 다른 수의 RAC 노드를 사용하는 것입니다. 이 경우 3차 사이트 Tiebreaker를 사용하는 것이 좋습니다. 세 번째 사이트를 사용할 수 없는 경우 Tiebreaker 인스턴스를 두 사이트의 한 사이트에 배치하여 해당 사이트를 효율적으로 기본 생존자 사이트로 지정할 수 있습니다.

Oracle 및 `css_critical`

노드 수가 짝수일 경우 기본 Oracle RAC 동작은 클러스터의 노드 중 하나가 다른 노드보다 더 중요한 것으로 간주된다는 것입니다. 우선 순위가 더 높은 노드가 있는 사이트는 사이트 격리를 유지하는 반면 다른 사이트의 노드는 제거됩니다. 우선 순위는 여러 요소를 기반으로 하지만 설정을 사용하여 이 동작을 제어할 수도 `css_critical` 있습니다.

"예" 아키텍처에서 RAC 노드의 호스트 이름은 `jfs12` 및 `jfs13`입니다. 의 현재 설정은 `css_critical` 다음과 같습니다.

```
[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.
```

`jfs12`가 있는 사이트를 기본 사이트로 설정하려면 사이트 A 노드에서 이 값을 예로 변경하고 서비스를 다시 시작합니다.

```
[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.
```

실패 시나리오

개요

전체 SnapMirror 액티브 동기화 애플리케이션 아키텍처를 계획하려면 계획된 페일오버 및 예상치 못한 다양한 페일오버 시나리오에서 SM-AS가 어떻게 반응하는지 이해해야 합니다.

다음 예에서는 사이트 A가 기본 사이트로 구성되어 있다고 가정합니다.

복제 접속이 끊어졌습니다

SM-AS 복제가 중단되면 클러스터에서 변경 내용을 반대편 사이트로 복제할 수 없기 때문에 쓰기 입출력을 완료할 수 없습니다.

사이트 A(기본 사이트)

기본 사이트에서 복제 링크 실패의 결과는 ONTAP가 복제된 쓰기 작업을 다시 시도하기 때문에 쓰기 입출력 처리가 약 15초 동안 일시 중지되는 것입니다. 이 경우 복제 링크에 도달할 수 없는 것으로 판단됩니다. 15초가 지나면 사이트 A 시스템이 읽기 및 쓰기 IO 처리를 재개합니다. SAN 경로는 변경되지 않으며 LUN은 온라인 상태로 유지됩니다.

사이트 B

사이트 B는 SnapMirror 활성 동기화 기본 사이트가 아니므로 약 15초 후에 해당 LUN 경로를 사용할 수 없게 됩니다.

스토리지 시스템 장애

스토리지 시스템 장애의 결과는 복제 링크 손실의 결과와 거의 동일합니다. 정상적인 사이트에서 약 15초의 입출력 일시 중지 시간이 발생합니다. 15초가 지나면 평소와 같이 해당 사이트에서 입출력이 재개됩니다.

중재자의 상실

중재자 서비스는 스토리지 운영을 직접 제어하지 않습니다. 클러스터 간 대체 제어 경로 역할을 합니다. 이는 주로 브레인 분할 시나리오의 위험 없이 장애 조치를 자동화하는 데 있습니다. 정상 작동 시 각 클러스터가 파트너에 변경 사항을 복제하고 있으므로 각 클러스터가 온라인 상태이고 데이터를 제공하고 있는지 확인할 수 있습니다. 복제 링크가 실패하면 복제가 중지됩니다.

안전한 자동 페일오버를 위해 중재자가 필요한 이유는 스토리지 클러스터에서 양방향 통신 손실이 네트워크 중단이나 실제 스토리지 장애로 인한 것인지 여부를 확인할 수 없기 때문입니다.

중재자는 각 클러스터에서 파트너 상태를 확인할 수 있는 대체 경로를 제공합니다. 시나리오는 다음과 같습니다.

- 클러스터가 파트너에게 직접 연락할 수 있는 경우 복제 서비스가 작동합니다. 별도의 조치가 필요 없습니다.
- 기본 사이트가 파트너에게 직접 연락하거나 중재자를 통해 연락할 수 없는 경우, 해당 파트너가 실제로 사용할 수 없거나 격리되어 해당 LUN 경로를 오프라인으로 설정한 것으로 간주됩니다. 그러면 기본 사이트가 RPO=0 상태를 해제하고 읽기 및 쓰기 입출력을 계속 처리합니다.
- 비선호 사이트가 해당 파트너에 직접 연락할 수 없지만 중재자를 통해 연락할 수 있는 경우 해당 경로가 오프라인 상태가 되고 복제 연결이 반환될 때까지 기다립니다.
- 비선호 사이트가 파트너에게 직접 연락하거나 운영 중재자를 통해 연락할 수 없는 경우, 파트너는 실제로 파트너를 사용할 수 없거나 격리되어 LUN 경로를 오프라인으로 전환했다고 가정합니다. 그러면 비기본 사이트가 RPO=0 상태를 해제하고 읽기 및 쓰기 입출력을 계속 처리합니다. 복제 소스의 역할을 가정하고 새로운 기본 사이트가 됩니다.

중재자를 완전히 사용할 수 없는 경우:

- 기본 설정되지 않은 사이트 또는 스토리지 시스템의 장애를 포함하여 어떠한 이유로든 복제 서비스에 장애가 발생하면 기본 사이트에서 RPO=0 상태를 해제하고 읽기 및 쓰기 입출력 처리를 재개합니다. 기본 사이트가 아닌 사이트는 해당 경로를 오프라인으로 전환합니다.
- 기본 사이트에 장애가 발생하면 기본 사이트가 반대 사이트가 실제로 오프라인 상태인지 확인할 수 없으므로 기본 사이트가 서비스를 다시 시작하는 것이 안전하지 않으므로 운영 중단이 발생합니다.

서비스를 복원하는 중입니다

사이트 간 연결 복원 또는 장애 시스템의 전원 켜기와 같은 장애가 해결되면 SnapMirror 활성 동기화 엔드포인트는 장애가 있는 복제 관계의 존재를 자동으로 감지하여 RPO=0 상태로 되돌립니다. 동기식 복제가 다시 설정되면 장애가 발생한 경로가 다시 온라인 상태가 됩니다.

대부분의 경우, 클러스터된 애플리케이션은 장애가 발생한 경로의 반환을 자동으로 감지하여 다시 온라인 상태로 돌아갑니다. 호스트 레벨 SAN 검사가 필요하거나 애플리케이션을 수동으로 다시 온라인으로 전환해야 하는 경우도 있습니다. 애플리케이션 및 구성 방법에 따라 다르며 일반적으로 이러한 작업을 쉽게 자동화할 수 있습니다. ONTAP 자체는 자동 복구이므로 RPO=0 스토리지 작업을 재개하기 위해 사용자 개입이 필요하지 않습니다.

수동 페일오버

기본 사이트를 변경하려면 간단한 작업이 필요합니다. 클러스터 간 복제 동작 전환에 대한 권한으로 입출력이 1-2초 동안 일시 중지되지만, 그렇지 않으면 입출력이 영향을 받지 않습니다.

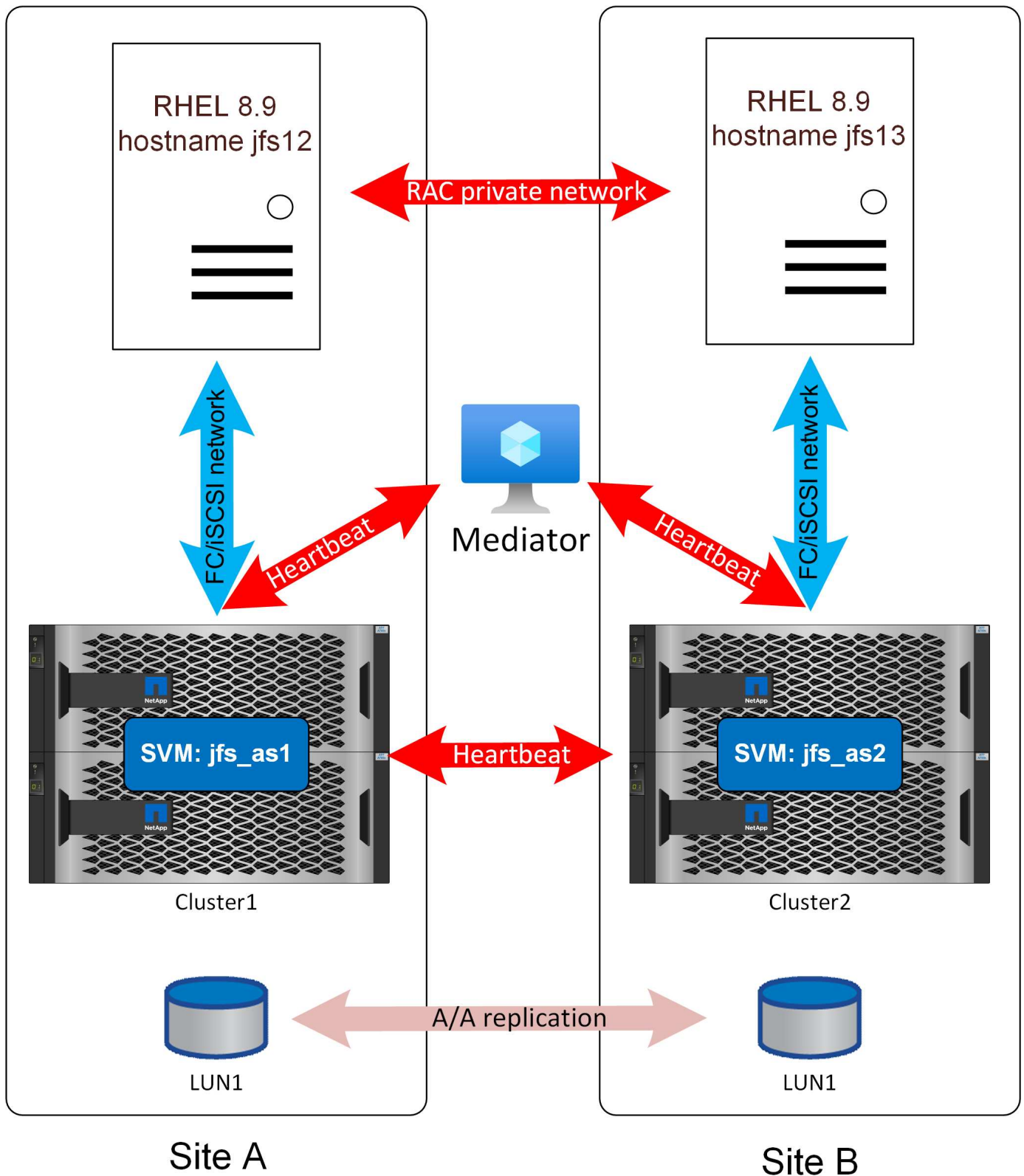
아키텍처 예

이 섹션에 나와 있는 자세한 장애 예는 아래에 나와 있는 아키텍처를 기반으로 합니다.



이는 SnapMirror 액티브 동기화에서 Oracle 데이터베이스를 위한 여러 옵션 중 하나일 뿐입니다. 이 설계는 보다 복잡한 시나리오를 보여 주기 때문에 선택되었습니다.

이 디자인에서는 사이트 A가 에 설정되어 있다고 "기본 사이트"가정합니다.



RAC 상호 연결 오류입니다

Oracle RAC 복제 링크가 손실되면 SnapMirror 접속 손실과 유사한 결과가 발생합니다. 단, 기본적으로 시간 초과가 짧아진다는 점이 다릅니다. 기본 설정에서 Oracle RAC 노드는 스토리지 연결이 끊긴 후 200초 동안 대기했다가 꺼내지만 RAC 네트워크 하트비트가 손실된 후

30초만 대기합니다.

CRS 메시지는 아래와 유사합니다. 30초의 타임아웃 타임랩스를 볼 수 있습니다. 사이트 A에 있는 jfs12 jfss_critical이 설정되어 있기 때문에 사이트 B에 있는 jfs13이 제거됩니다.

```
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 통신 실패

SnapMirror 활성화 동기화 복제 링크인 경우 클러스터에서 변경 내용을 반대편 사이트로 복제할 수 없기 때문에 쓰기 입출력을 완료할 수 없습니다.

사이트 A

복제 링크가 제대로 작동하지 않는 것으로 판단되기 전에 ONTAP에서 쓰기 복제를 시도하기 때문에 사이트 A에서 쓰기 입출력 처리가 약 15초 동안 일시 중지됩니다. 15초가 지나면 사이트 A의 ONTAP 클러스터가 읽기 및 쓰기 IO 처리를 재개합니다. SAN 경로는 변경되지 않으며 LUN은 온라인 상태로 유지됩니다.

사이트 B

사이트 B는 SnapMirror 활성화 동기화 기본 사이트가 아니므로 약 15초 후에 해당 LUN 경로를 사용할 수 없게 됩니다.

복제 링크가 타임 스탬프 15:19:44에서 잘렸습니다. Oracle RAC의 첫 번째 경고는 200초 시간 초과(Oracle RAC 매개 변수 disktimeout에 의해 제어됨)가 접근함에 따라 100초 후에 도착합니다.

```

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.

```

200초의 보팅 디스크 시간 초과에 도달하면 이 Oracle RAC 노드가 클러스터에서 자체적으로 제거되고 재부팅됩니다.

전체 네트워크 상호 연결 실패

사이트 간 복제 링크가 완전히 손실되면 SnapMirror 활성 동기화와 Oracle RAC 연결이 모두 중단됩니다.

Oracle RAC의 브레인 분할 감지는 Oracle RAC 스토리지 하트비트에 의존합니다. 사이트 간 연결이 끊겨 RAC 네트워크 하트비트 및 스토리지 복제 서비스가 동시에 손실되면 RAC 사이트가 RAC 상호 연결 또는 RAC 보팅 디스크를 통해 교차 사이트와 통신할 수 없게 됩니다. 결과적으로 짝수 노드 집합이 기본 설정에서 두 사이트를 모두 제거할 수 있습니다. 정확한 동작은 이벤트의 순서와 RAC 네트워크 및 디스크 하트비트 폴링의 시간에 따라 달라집니다.

2개 사이트 중단 위험은 두 가지 방법으로 해결할 수 있습니다. 먼저, "[타이브레이커](#)" 설정을 사용할 수 있습니다.

세 번째 사이트를 사용할 수 없는 경우 RAC 클러스터에서 miscount 매개 변수를 조정하여 이 위험을 해결할 수 있습니다. 기본값에서 RAC 네트워크 하트비트 시간 초과는 30초입니다. 일반적으로 RAC는 장애가 발생한 RAC 노드를 식별하여 클러스터에서 제거하는 데 사용됩니다. 또한 보팅 디스크 하트비트에도 연결되어 있습니다.

예를 들어, Oracle RAC와 스토리지 복제 서비스 모두에 대해 사이트 간 트래픽을 전달하는 도관이 백호를 통해 삭감되면 30초 오류 카운트다운이 시작됩니다. RAC 기본 사이트 노드가 30초 이내에 반대쪽 사이트와 연결을 다시 설정할 수 없고, 또한 투표 디스크를 사용하여 반대쪽 사이트가 동일한 30초 기간 내에 다운되었는지 확인할 수 없는 경우 기본 사이트 노드도 제거됩니다. 그 결과, 데이터베이스가 완전히 중단됩니다.

오류 카운트가 폴링되는 시점에 따라 30초가 부족하여 SnapMirror 활성 동기화 시간이 초과되고 30초 기간이 만료되기 전에 기본 사이트의 저장소가 서비스를 다시 시작할 수 없습니다. 이 30초 기간을 늘릴 수 있습니다.

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

이 값을 사용하면 기본 사이트의 스토리지 시스템에서 카운트 타임 아웃이 만료되기 전에 작업을 재개할 수 있습니다. 그러면 LUN 경로가 제거된 사이트에서 노드만 제거됩니다. 아래 예:

```
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
```

Oracle Support에서는 구성 문제를 해결하기 위해 misscount 또는 disktimeout 매개 변수를 변경하지 않는 것이 좋습니다. 그러나 이러한 매개 변수를 변경하는 것은 SAN 부팅, 가상화 및 스토리지 복제 구성을 포함하여 많은 경우에 불가피하게 수행할 수 있습니다. 예를 들어, SAN 또는 IP 네트워크에서 RAC 제거를 초래하는 안정성 문제가 발생한 경우 기본 문제를 해결하고 오류 카운트나 디스크 시간 초과 값을 청구하지 않아야 합니다. 구성 오류를 해결하기 위해 시간 제한을 변경하면 문제가 해결되는 것이 아니라 문제를 마스킹하는 것입니다. 기반 인프라의 설계 측면을 기반으로 RAC 환경을 올바르게 구성하기 위해 이러한 매개 변수를 변경하는 것은 Oracle 지원 명령어와 일관적입니다. SAN 부팅 시 디스크 시간 제한을 맞추기 위해 최대 200까지 misscount를 조정하는 것이 일반적입니다. 자세한 내용은 [이 링크](#)를 참조하십시오.

사이트 장애

스토리지 시스템 또는 사이트 장애 결과는 복제 링크 손실의 결과와 거의 동일합니다. 정상적인 사이트에서 쓰기 작업 시 약 15초의 입출력 일시 중지 현상이 발생합니다. 15초가 지나면 평소와

같이 해당 사이트에서 입출력이 재개됩니다.

스토리지 시스템만 영향을 받은 경우 장애가 발생한 사이트의 Oracle RAC 노드는 스토리지 서비스를 잃게 되고 제거 및 후속 재부팅 전에 동일한 200초 디스크 시간 초과 카운트다운을 시작합니다.

```
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.
```

스토리지 서비스가 손실된 RAC 노드의 SAN 경로 상태는 다음과 같습니다.

```
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
```

Linux 호스트가 200초보다 훨씬 빠른 경로 손실을 감지했지만 데이터베이스 관점에서 장애가 발생한 사이트의 호스트에 대한 클라이언트 연결은 기본 Oracle RAC 설정에서 200초 동안 동결됩니다. 전체 데이터베이스 작업은 제거가 완료된 후에만 재개됩니다.

한편, 반대쪽 사이트의 Oracle RAC 노드는 다른 RAC 노드의 손실을 기록합니다. 그렇지 않으면 평소와 같이 계속 작동합니다.

```
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
```

중재자 고장

중재자 서비스는 스토리지 운영을 직접 제어하지 않습니다. 클러스터 간 대체 제어 경로 역할을 합니다. 이는 주로 브레인 분할 시나리오의 위험 없이 장애 조치를 자동화하는 데 있습니다.

정상 작동 시 각 클러스터가 파트너에 변경 사항을 복제하고 있으므로 각 클러스터가 온라인 상태이고 데이터를 제공하고 있는지 확인할 수 있습니다. 복제 링크가 실패하면 복제가 중지됩니다.

안전한 자동 운영을 위해 중재자가 필요한 이유는 스토리지 클러스터에서 양방향 통신 손실이 네트워크 중단이나 실제 스토리지 장애로 인한 것인지 여부를 확인할 수 없기 때문입니다.

중재자는 각 클러스터에서 파트너 상태를 확인할 수 있는 대체 경로를 제공합니다. 시나리오는 다음과 같습니다.

- 클러스터가 파트너에게 직접 연락할 수 있는 경우 복제 서비스가 작동합니다. 별도의 조치가 필요 없습니다.
- 기본 사이트가 파트너에게 직접 연락하거나 중재자를 통해 연락할 수 없는 경우, 해당 파트너가 실제로 사용할 수 없거나 격리되어 해당 LUN 경로를 오프라인으로 설정한 것으로 간주됩니다. 그러면 기본 사이트가 RPO=0 상태를 해제하고 읽기 및 쓰기 입출력을 계속 처리합니다.
- 비선호 사이트가 해당 파트너에 직접 연락할 수 없지만 중재자를 통해 연락할 수 있는 경우 해당 경로가 오프라인 상태가 되고 복제 연결이 반환될 때까지 기다립니다.
- 비선호 사이트가 파트너에게 직접 연락하거나 운영 중재자를 통해 연락할 수 없는 경우, 파트너는 실제로 파트너를

사용할 수 없거나 격리되어 LUN 경로를 오프라인으로 전환했다고 가정합니다. 그러면 비기본 사이트가 RPO=0 상태를 해제하고 읽기 및 쓰기 입출력을 계속 처리합니다. 복제 소스의 역할을 가정하고 새로운 기본 사이트가 됩니다.

중재자를 완전히 사용할 수 없는 경우:

- 어떤 이유로든 복제 서비스가 실패하면 기본 사이트가 RPO=0 상태를 해제하고 읽기 및 쓰기 입출력 처리를 재개합니다. 기본 사이트가 아닌 사이트는 해당 경로를 오프라인으로 전환합니다.
- 기본 사이트에 장애가 발생하면 기본 사이트가 반대 사이트가 실제로 오프라인 상태인지 확인할 수 없으므로 기본 사이트가 서비스를 다시 시작하는 것이 안전하지 않으므로 운영 중단이 발생합니다.

서비스 복원

SnapMirror는 자가 복구입니다. SnapMirror 활성 동기화는 장애가 있는 복제 관계가 있는지 자동으로 감지하여 RPO=0 상태로 되돌립니다. 동기식 복제가 다시 설정되면 경로가 다시 온라인 상태가 됩니다.

대부분의 경우, 클러스터된 애플리케이션은 장애가 발생한 경로의 반환을 자동으로 감지하여 다시 온라인 상태로 돌아갑니다. 호스트 레벨 SAN 검사가 필요하거나 애플리케이션을 수동으로 다시 온라인으로 전환해야 하는 경우도 있습니다.

애플리케이션 및 구성 방식에 따라 다르며 일반적으로 이러한 작업을 쉽게 자동화할 수 있습니다. SnapMirror 활성 동기화 자체는 자체 해결 방식이므로 전원 및 연결이 복원된 후 RPO=0 스토리지 작업을 재개하기 위해 사용자 개입이 필요하지 않습니다.

수동 페일오버

"페일오버"라는 용어는 양방향 복제 기술이기 때문에 SnapMirror 활성 동기화를 사용하는 복제 방향을 가리키지 않습니다. 대신 '페일오버'는 장애 발생 시 기본 사이트가 될 스토리지 시스템을 나타냅니다.

예를 들어, 유지보수를 위해 사이트를 종료하기 전에 또는 DR 테스트를 수행하기 전에 페일오버를 수행하여 기본 사이트를 변경할 수 있습니다.

기본 사이트를 변경하려면 간단한 작업이 필요합니다. 클러스터 간 복제 동작 전환에 대한 권한으로 입출력이 1-2초 동안 일시 중지되지만, 그렇지 않으면 입출력이 영향을 받지 않습니다.

GUI 예:

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>		

CLI를 통해 다시 변경하는 예:

```
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
```

저작권 정보

Copyright © 2026 NetApp, Inc. All Rights Reserved. 미국에서 인쇄된 본 문서의 어떠한 부분도 저작권 소유자의 사전 서면 승인 없이는 어떠한 형식이나 수단(복사, 녹음, 녹화 또는 전자 검색 시스템에 저장하는 것을 비롯한 그래픽, 전자적 또는 기계적 방법)으로도 복제될 수 없습니다.

NetApp이 저작권을 가진 자료에 있는 소프트웨어에는 아래의 라이선스와 고지사항이 적용됩니다.

본 소프트웨어는 NetApp에 의해 '있는 그대로' 제공되며 상품성 및 특정 목적에의 적합성에 대한 명시적 또는 묵시적 보증을 포함하여(이에 제한되지 않음) 어떠한 보증도 하지 않습니다. NetApp은 대체품 또는 대체 서비스의 조달, 사용 불능, 데이터 손실, 이익 손실, 영업 중단을 포함하여(이에 국한되지 않음), 이 소프트웨어의 사용으로 인해 발생하는 모든 직접 및 간접 손해, 우발적 손해, 특별 손해, 징벌적 손해, 결과적 손해의 발생에 대하여 그 발생 이유, 책임론, 계약 여부, 엄격한 책임, 불법 행위(과실 또는 그렇지 않은 경우)와 관계없이 어떠한 책임도 지지 않으며, 이와 같은 손실의 발생 가능성이 통지되었다 하더라도 마찬가지입니다.

NetApp은 본 문서에 설명된 제품을 언제든지 예고 없이 변경할 권리를 보유합니다. NetApp은 NetApp의 명시적인 서면 동의를 받은 경우를 제외하고 본 문서에 설명된 제품을 사용하여 발생하는 어떠한 문제에도 책임을 지지 않습니다. 본 제품의 사용 또는 구매의 경우 NetApp에서는 어떠한 특허권, 상표권 또는 기타 지적 재산권이 적용되는 라이선스도 제공하지 않습니다.

본 설명서에 설명된 제품은 하나 이상의 미국 특허, 해외 특허 또는 출원 중인 특허로 보호됩니다.

제한적 권리 표시: 정부에 의한 사용, 복제 또는 공개에는 DFARS 252.227-7013(2014년 2월) 및 FAR 52.227-19(2007년 12월)의 기술 데이터-비상업적 품목에 대한 권리(Rights in Technical Data -Noncommercial Items) 조항의 하위 조항 (b)(3)에 설명된 제한사항이 적용됩니다.

여기에 포함된 데이터는 상업용 제품 및/또는 상업용 서비스(FAR 2.101에 정의)에 해당하며 NetApp, Inc.의 독점 자산입니다. 본 계약에 따라 제공되는 모든 NetApp 기술 데이터 및 컴퓨터 소프트웨어는 본질적으로 상업용이며 개인 비용만으로 개발되었습니다. 미국 정부는 데이터가 제공된 미국 계약과 관련하여 해당 계약을 지원하는 데에만 데이터에 대한 전 세계적으로 비독점적이고 양도할 수 없으며 재사용이 불가능하며 취소 불가능한 라이선스를 제한적으로 가집니다. 여기에 제공된 경우를 제외하고 NetApp, Inc.의 사전 서면 승인 없이는 이 데이터를 사용, 공개, 재생산, 수정, 수행 또는 표시할 수 없습니다. 미국 국방부에 대한 정부 라이선스는 DFARS 조항 252.227-7015(b)(2014년 2월)에 명시된 권한으로 제한됩니다.

상표 정보

NETAPP, NETAPP 로고 및 <http://www.netapp.com/TM>에 나열된 마크는 NetApp, Inc.의 상표입니다. 기타 회사 및 제품 이름은 해당 소유자의 상표일 수 있습니다.