



Synchronisation active SnapMirror

Enterprise applications

NetApp
January 02, 2026

Sommaire

Synchronisation active SnapMirror.....	1
Présentation	1
Réplication synchrone	1
Matériel de stockage	1
ONTAP Médiateur	1
Médiateur de ONTAP	1
Site préféré de la synchronisation active SnapMirror.....	3
Topologie réseau	4
Accès uniforme	4
Accès non uniforme	9
Configurations Oracle.....	11
Présentation	11
Instance unique Oracle	11
RAC étendu Oracle	13
Disjoncteur d'attache RAC	22
Scénarios d'échec	23
Présentation	23
Exemple d'architecture.....	25
Échec de l'interconnexion du RAC	26
Échec de communication SnapMirror	27
Échec total de l'interconnectivité réseau	28
Panne du site	30
Défaillance du médiateur	32
Restauration du service	33
Basculement manuel	33

Synchronisation active SnapMirror

Présentation

La synchronisation active SnapMirror vous permet de créer des environnements de base de données Oracle à ultra haute disponibilité où des LUN sont disponibles à partir de deux clusters de stockage différents.

Avec la synchronisation active SnapMirror, il n'y a pas de copie « principale » ni de copie « secondaire » des données. Chaque cluster peut fournir des E/S de lecture à partir de sa copie locale des données, et chaque cluster réplique une écriture vers son partenaire. Le résultat est un comportement d'E/S symétrique.

Entre autres options, vous pouvez exécuter Oracle RAC en tant que cluster étendu avec des instances opérationnelles sur les deux sites. Vous pouvez également créer des clusters de bases de données actif-passif RPO=0, dans lesquels les bases de données à instance unique peuvent être déplacées entre les sites en cas de panne sur le site. Ce processus peut également être automatisé via des produits tels que Pacemaker ou VMware HA. Toutes ces options reposent sur la réPLICATION synchrone gérée par SnapMirror Active Sync.

RéPLICATION synchrone

En fonctionnement normal, la synchronisation active SnapMirror fournit en permanence une réPLIQUE synchrone avec un objectif de point de récupération de 0, à une exception près. Si les données ne peuvent pas être répliquées, ONTAP exige de répliquer les données et de reprendre le traitement des E/S sur un site pendant que les LUN de l'autre site sont mises hors ligne.

Matériel de stockage

Contrairement à d'autres solutions de reprise après incident du stockage, la synchronisation active SnapMirror offre une flexibilité asymétrique de la plateforme. Le matériel de chaque site n'a pas besoin d'être identique. Cette fonctionnalité vous permet d'ajuster la taille du matériel utilisé pour prendre en charge la synchronisation active SnapMirror. Le système de stockage distant peut être identique au site principal s'il doit prendre en charge une charge de travail de production complète, mais si un incident entraîne une réduction des E/S, un système plus petit sur le site distant peut être plus économique.

ONTAP Médiateur

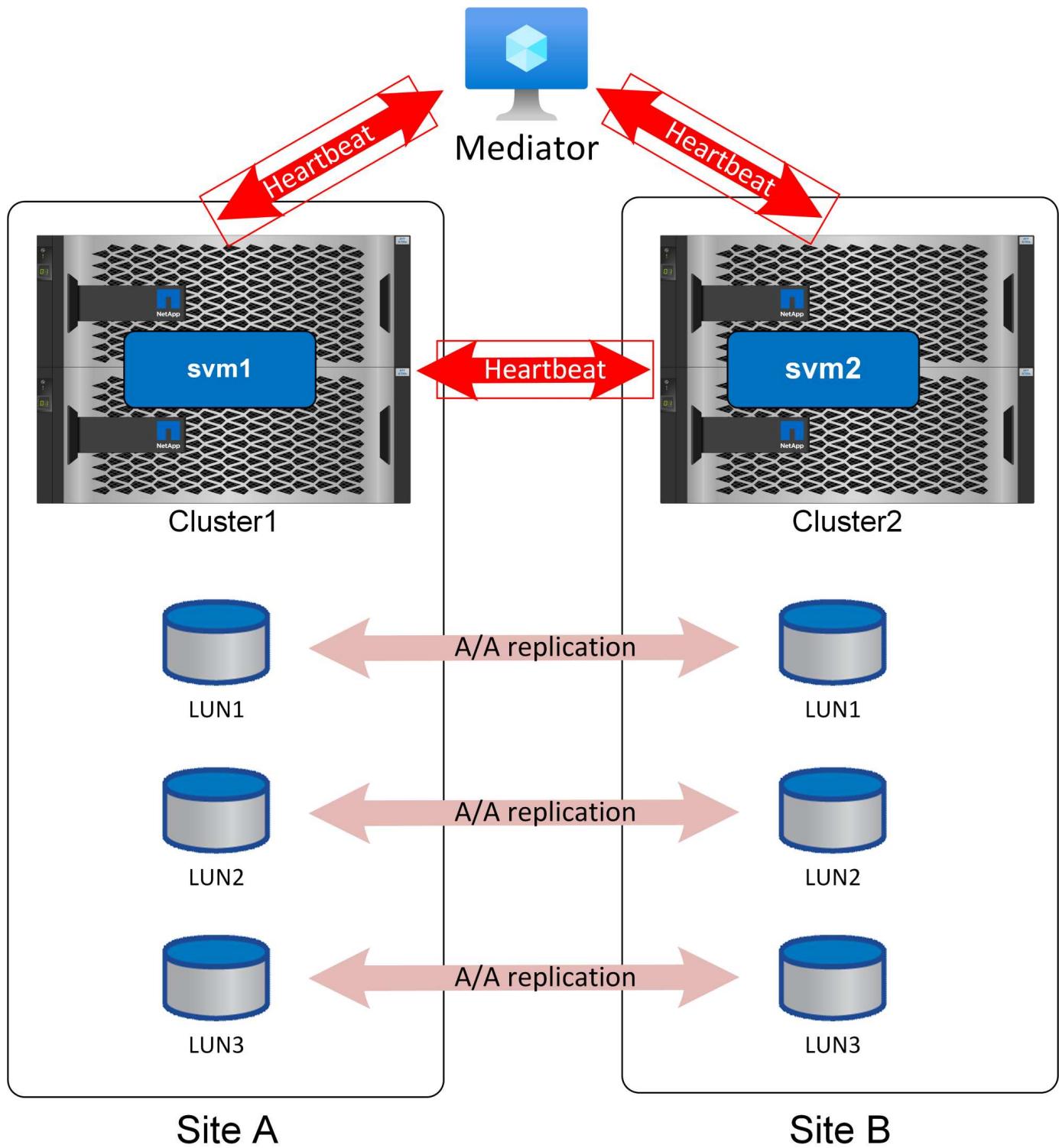
Le médiateur ONTAP est une application logicielle téléchargée depuis la prise en charge de NetApp et généralement déployée sur une petite machine virtuelle. Le Mediator ONTAP n'est pas un Tiebreaker lorsqu'il est utilisé avec la synchronisation active SnapMirror. Il s'agit d'un canal de communication alternatif pour les deux clusters qui participent à la réPLICATION SnapMirror active Sync. Les opérations automatisées sont dirigées par ONTAP sur la base des réponses reçues du partenaire via des relations directes et via le médiateur.

Médiateur de ONTAP

Le médiateur est requis pour automatiser le basculement en toute sécurité. Dans l'idéal, elle serait placée sur un site tiers indépendant, mais elle peut toujours fonctionner pour la plupart des besoins si elle est en colocation avec l'un des clusters participant à la réPLICATION.

Le médiateur n'est pas vraiment un arbitre, même si c'est effectivement la fonction qu'il remplit. Le médiateur

aide à déterminer l'état des nœuds du cluster et facilite le processus de basculement automatique en cas de panne d'un site. Le médiateur ne transfère aucune donnée, quelles que soient les circonstances.



Le principal défi lié au basculement automatisé est le problème des réseaux partagés, qui se pose en cas de perte de connectivité entre les deux sites. Que doit-on faire ? Vous ne voulez pas que deux sites différents se désignent comme les copies restantes des données, mais comment un seul site peut-il faire la différence entre la perte réelle du site opposé et l'incapacité à communiquer avec le site opposé ?

C'est là que le médiateur entre dans la photo. S'il est placé sur un troisième site, et chaque site a une

connexion réseau distincte à ce site, alors vous avez un chemin supplémentaire pour chaque site pour valider l'état de santé de l'autre. Examinez à nouveau l'image ci-dessus et examinez les scénarios suivants.

- Que se passe-t-il si le médiateur échoue ou est inaccessible à partir d'un ou des deux sites ?
 - Les deux clusters peuvent toujours communiquer entre eux sur le même lien que celui utilisé pour les services de réPLICATION.
 - Les données restent protégées avec un objectif de point de récupération de 0
- Que se passe-t-il si le site A tombe en panne ?
 - Le site B verra les deux canaux de communication tomber en panne.
 - Le site B prendra le contrôle des services de données, mais sans mise en miroir RPO=0
- Que se passe-t-il si le site B tombe en panne ?
 - Le site A verra les deux canaux de communication tomber en panne.
 - Le site A prend le relais des services de données, mais sans mise en miroir avec un objectif de point de récupération de 0

Il y a un autre scénario à prendre en compte : la perte du lien de réPLICATION des données. En cas de perte de la liaison de réPLICATION entre les sites, la mise en miroir avec un objectif de point de récupération de 0 sera évidemment impossible. Que devrait-on alors se passer ?

Ceci est contrôlé par le statut du site préféré. Dans une relation SM-AS, l'un des sites est secondaire à l'autre. Cela n'a aucun effet sur les opérations normales, et tout accès aux données est symétrique. Toutefois, si la réPLICATION est interrompue, le nœud devra être rompu pour reprendre les opérations. Par conséquent, le site privilégié continuera les opérations sans mise en miroir et le site secondaire arrêtera le traitement des E/S jusqu'à ce que la communication de réPLICATION soit restaurée.

Site préféré de la synchronisation active SnapMirror

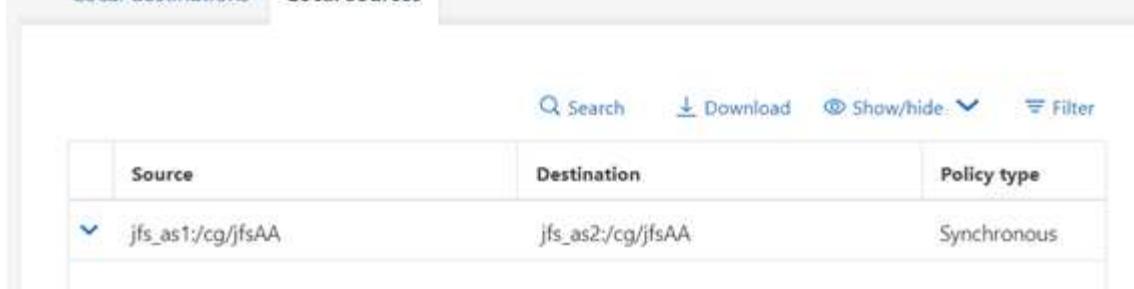
Le comportement de la synchronisation active SnapMirror est symétrique, avec une exception importante : la configuration du site préféré.

La synchronisation active SnapMirror considère un site comme la « source » et l'autre comme la « destination ». Cela implique une relation de réPLICATION unidirectionnelle, mais cela ne s'applique pas au comportement d'E/S. La réPLICATION est bidirectionnelle et symétrique. Les temps de réponse d'E/S sont identiques de part et d'autre du miroir.

La source désignation est le contrôle du site préféré. En cas de perte du lien de réPLICATION, les chemins de LUN sur la copie source continueront à transmettre des données tandis que les chemins de LUN sur la copie de destination deviendront indisponibles jusqu'à ce que la réPLICATION soit rétablie et que SnapMirror repasse à l'état synchrone. Les chemins reprennent alors le service des données.

La configuration source/destination peut être affichée via SystemManager :

Relationships



Source	Destination	Policy type
jfs_as1:/cg/jfsAA	jfs_as2:/cg/jfsAA	Synchronous

Ou sur l'interface de ligne de commande :

```
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
```

La clé est que la source est le SVM sur le cluster1. Comme mentionné ci-dessus, les termes « source » et « destination » ne décrivent pas le flux des données répliquées. Les deux sites peuvent traiter une écriture et la répliquer sur le site opposé. En effet, les deux grappes sont des sources et des destinations. La désignation d'un cluster comme source contrôle simplement le cluster qui survit en tant que système de stockage en lecture/écriture en cas de perte du lien de réPLICATION.

Topologie réseau

Accès uniforme

Un réseau d'accès uniforme signifie que les hôtes peuvent accéder aux chemins sur les deux sites (ou domaines de défaillance au sein du même site).

L'une des caractéristiques importantes de SM-AS est la capacité de configurer les systèmes de stockage pour savoir où se trouvent les hôtes. Lorsque vous mappez les LUN sur un hôte donné, vous pouvez indiquer si elles sont proximales ou non à un système de stockage donné.

Paramètres de proximité

La proximité fait référence à une configuration par cluster qui indique qu'un WWN d'hôte ou un ID d'initiateur

iSCSI appartient à un hôte local. Il s'agit d'une deuxième étape facultative de configuration de l'accès aux LUN.

La première étape correspond à la configuration habituelle du groupe initiateur. Chaque LUN doit être mappée sur un groupe initiateur qui contient les ID WWN/iSCSI des hôtes devant accéder à cette LUN. Cela contrôle quel hôte a accès à un LUN.

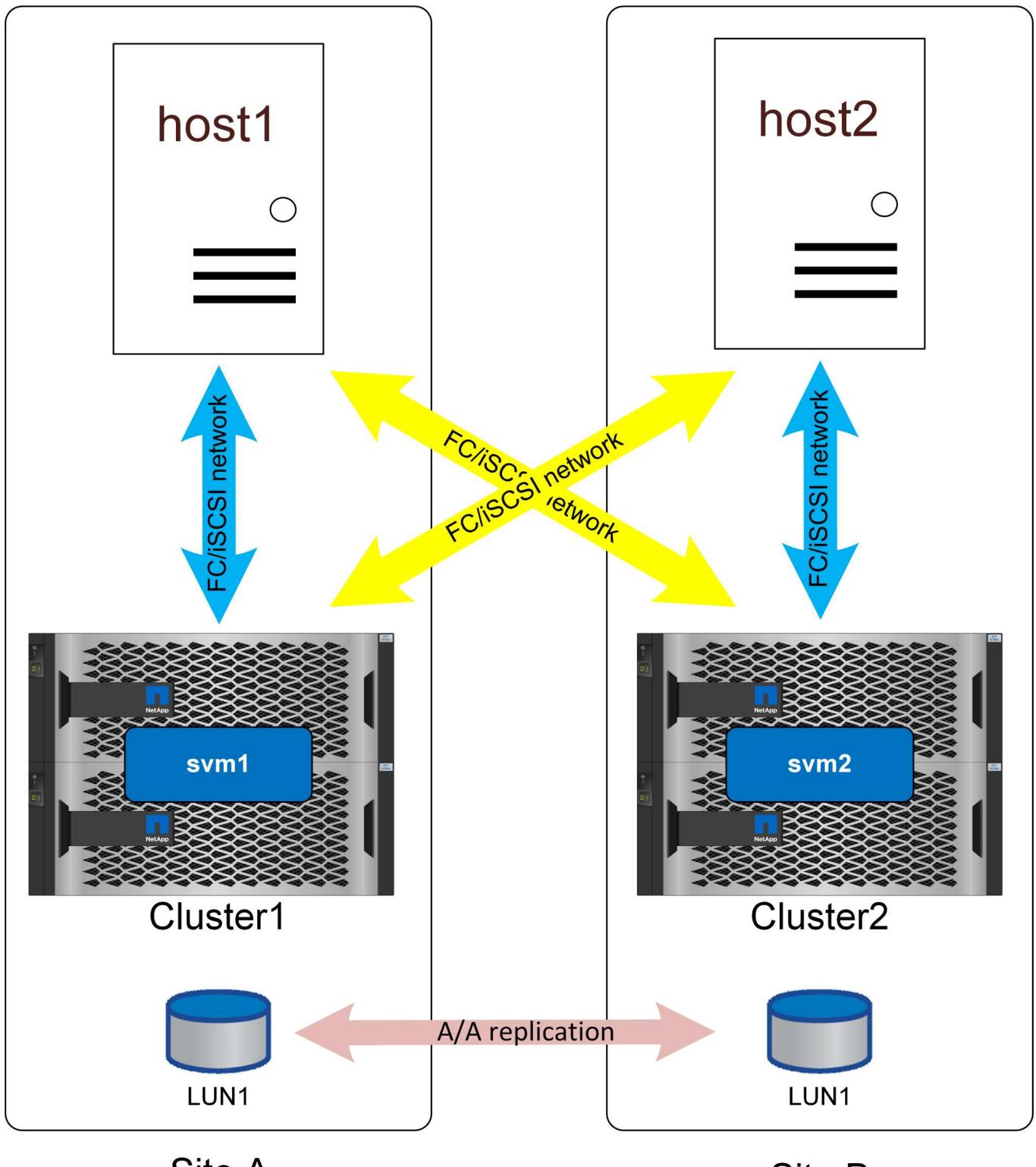
La deuxième étape facultative consiste à configurer la proximité de l'hôte. Cela ne contrôle pas l'accès, il contrôle *Priority*.

Par exemple, un hôte du site A peut être configuré pour accéder à une LUN protégée par la synchronisation active SnapMirror. Le SAN étant étendu entre les sites, les chemins d'accès sont disponibles pour cette LUN via le stockage sur le site A ou le stockage sur le site B.

Sans paramètres de proximité, cet hôte utilisera les deux systèmes de stockage de la même manière, car les deux systèmes de stockage annonceront des chemins actifs/optimisés. Si la latence SAN et/ou la bande passante entre les sites est limitée, il se peut que cela ne soit pas désirable, et vous pouvez vous assurer que, pendant le fonctionnement normal, chaque hôte utilise de préférence des chemins vers le système de stockage local. Cette configuration s'effectue en ajoutant l'ID WWN/iSCSI de l'hôte au cluster local en tant qu'hôte proximal. Cette opération peut être effectuée à partir de l'interface de ligne de commande ou de SystemManager.

AFF

Avec un système AFF, les chemins apparaissent comme indiqué ci-dessous lorsque la proximité de l'hôte a été configurée.



Active/Optimized Path

Active Path

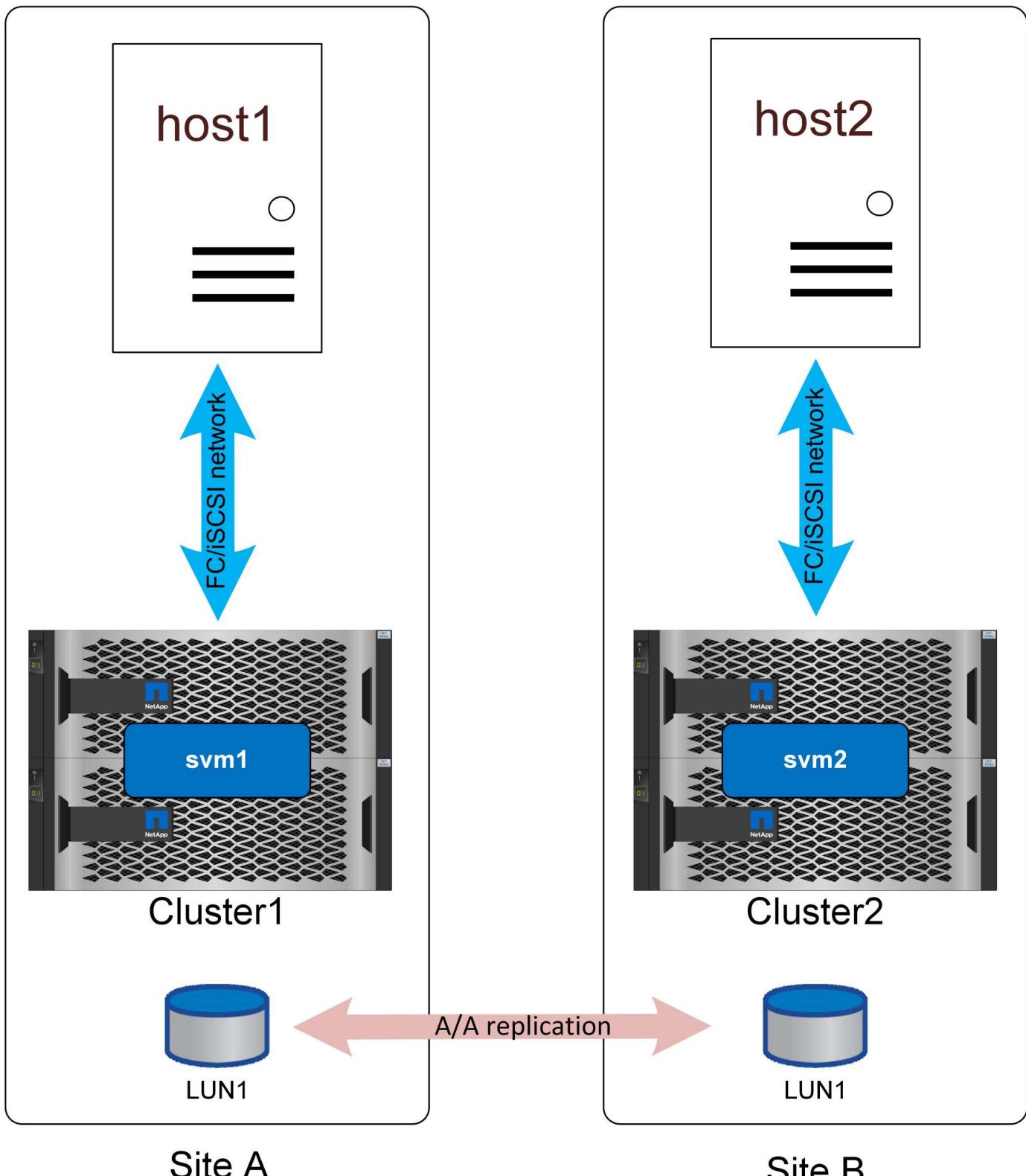
En fonctionnement normal, toutes les E/S sont des E/S locales. Les opérations de lecture et d'écriture sont gérées à partir de la baie de stockage locale. Bien entendu, les E/S en écriture devront également être répliquées par le contrôleur local sur le système distant avant d'être acquittées, mais toutes les E/S en lecture seront gérées localement et ne subiront pas de latence supplémentaire en traversant la liaison SAN entre les sites.

Le seul moment où les chemins non optimisés seront utilisés est la perte de tous les chemins actifs/optimisés. Par exemple, si l'ensemble de la baie sur le site A est hors tension, les hôtes du site A peuvent toujours accéder aux chemins d'accès à la baie sur le site B et donc rester opérationnels, même s'ils connaissent une latence plus élevée.

Il existe des chemins redondants à travers le cluster local qui ne sont pas illustrés sur ces schémas pour plus de simplicité. Les systèmes de stockage ONTAP étant dotés de la haute disponibilité, une panne du contrôleur ne devrait pas entraîner de panne sur le site. Il devrait simplement entraîner une modification dans laquelle les chemins locaux sont utilisés sur le site affecté.

ASA

Les systèmes NetApp ASA proposent des chemins d'accès multiples actif-actif sur tous les chemins d'accès à un cluster. Cela s'applique également aux configurations SM-AS.



Active/Optimized Path

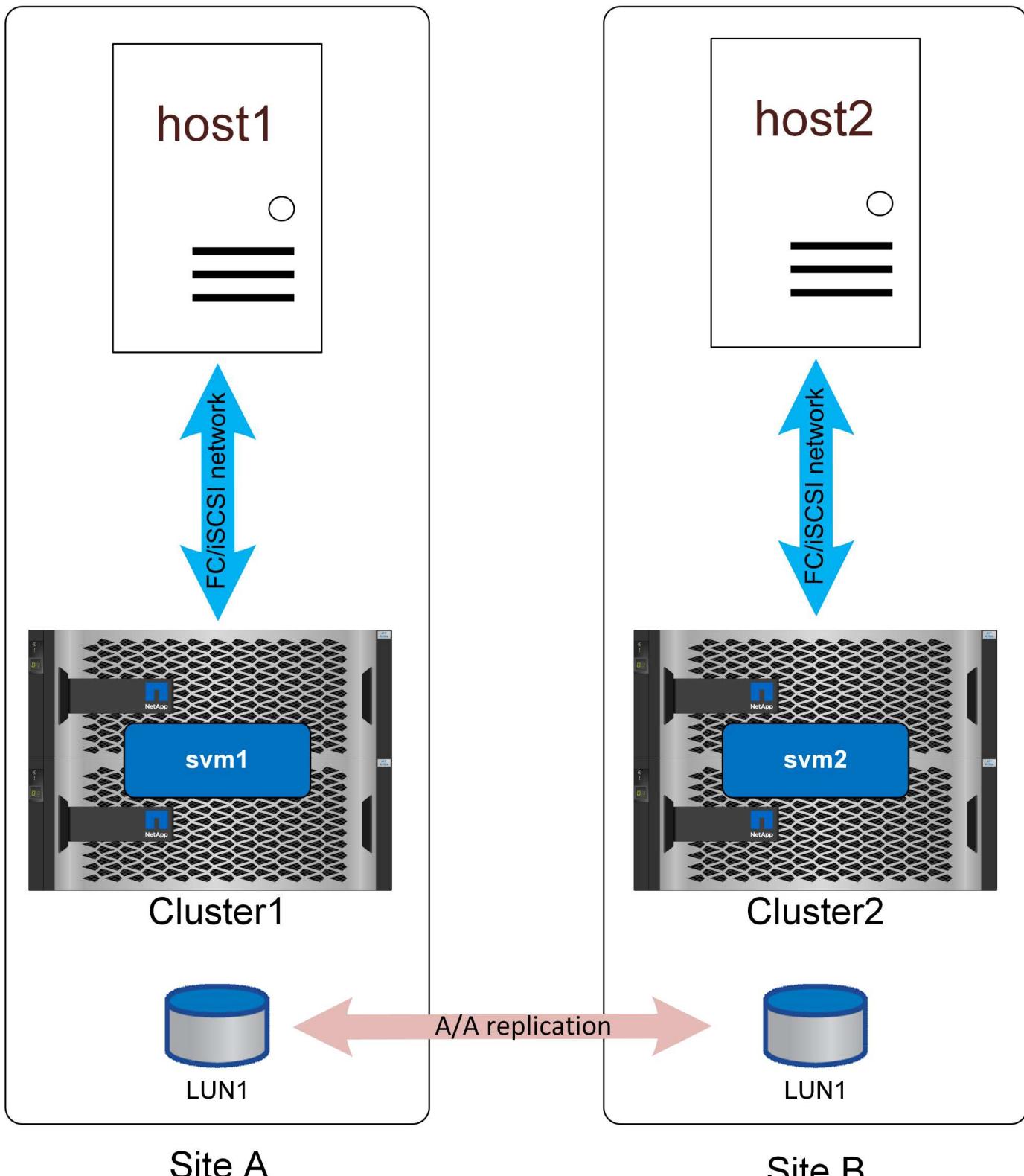
Une configuration ASA avec un accès non uniforme fonctionnera en grande partie comme avec AFF. Avec un accès uniforme, l'E/S traverserait le WAN. Cela peut être souhaitable ou non.

Si les deux sites étaient distants de 100 mètres avec une connectivité à fibre optique, il ne devrait pas y avoir de latence supplémentaire détectable traversant le WAN, mais si les sites étaient éloignés, les performances de lecture seraient affectées sur les deux sites. À l'inverse, avec AFF, ces chemins WAN seraient utilisés uniquement s'il n'existe aucun chemin local disponible et si les performances quotidiennes seraient meilleures, car toutes les E/S seraient des E/S locales. ASA avec un réseau d'accès non uniforme serait une option pour bénéficier des avantages de ASA en termes de coûts et de fonctionnalités sans engendrer de pénalités de latence entre les sites.

ASA avec SM-as dans une configuration à faible latence offre deux avantages intéressants. Tout d'abord, elle double *les performances de n'importe quel hôte, car les E/S peuvent être traitées par deux fois plus de contrôleurs en utilisant deux fois plus de chemins. Ensuite, dans un environnement à site unique, elle offre une disponibilité extrême, car l'intégralité du système de stockage peut être perdue sans interrompre l'accès aux hôtes.

Accès non uniforme

La mise en réseau à accès non uniforme signifie que chaque hôte n'a accès qu'aux ports du système de stockage local. Le SAN n'est pas étendu sur les sites (ou les domaines de défaillance au sein du même site).



Active/Optimized Path

Le principal avantage de cette approche est la simplicité du SAN : vous n'avez plus besoin d'étendre un SAN sur le réseau. Certains clients ne disposent pas d'une connectivité à faible latence suffisante entre les sites, ou

n'ont pas l'infrastructure nécessaire pour acheminer le trafic SAN FC sur un réseau intersite.

L'inconvénient de l'accès non uniforme est que certains scénarios de défaillance, notamment la perte du lien de réplication, entraînent la perte de l'accès au stockage par certains hôtes. En cas de perte de la connectivité du stockage local, les applications qui s'exécutent en tant qu'instances uniques, telles qu'une base de données non en cluster et qui ne s'exécute intrinsèquement que sur un hôte unique sur un montage donné, échouent. Les données seraient toujours protégées, mais le serveur de base de données n'aurait plus accès. Il doit être redémarré sur un site distant, de préférence par le biais d'un processus automatisé. Par exemple, VMware HA peut détecter une situation de tous les chemins d'accès sur un serveur et redémarrer une machine virtuelle sur un autre serveur sur lequel les chemins d'accès sont disponibles.

En revanche, une application en cluster telle qu'Oracle RAC peut fournir un service qui est disponible simultanément sur deux sites différents. La perte d'un site ne signifie pas la perte du service d'application dans son ensemble. Les instances restent disponibles et s'exécutent sur le site survivant.

Dans de nombreux cas, la surcharge liée à la latence supplémentaire qu'une application accède au système de stockage via une liaison site à site ne serait pas acceptable. Cela signifie que l'amélioration de la disponibilité des réseaux uniformes est minime, car la perte de stockage sur un site entraînerait la nécessité de fermer les services sur ce site défaillant.



Il existe des chemins redondants à travers le cluster local qui ne sont pas illustrés sur ces schémas pour plus de simplicité. Les systèmes de stockage ONTAP étant dotés de la haute disponibilité, une panne du contrôleur ne devrait pas entraîner de panne sur le site. Il devrait simplement entraîner une modification dans laquelle les chemins locaux sont utilisés sur le site affecté.

Configurations Oracle

Présentation

L'utilisation de la synchronisation active SnapMirror n'ajoute pas nécessairement aux meilleures pratiques d'exploitation d'une base de données ou ne modifie pas nécessairement ces pratiques.

La meilleure architecture dépend des besoins de l'entreprise. Par exemple, si l'objectif est de bénéficier d'une protection RPO=0 contre la perte de données, mais que l'objectif RTO est assoupli, l'utilisation de bases de données Oracle Single instance et la réplication des LUN avec SM-AS peuvent suffire et être moins coûteuses d'un standard de licences Oracle. Toute panne du site distant n'interrompt pas les opérations, et la perte du site principal entraînerait la présence de LUN en ligne et prêtes à être utilisées sur le site survivant.

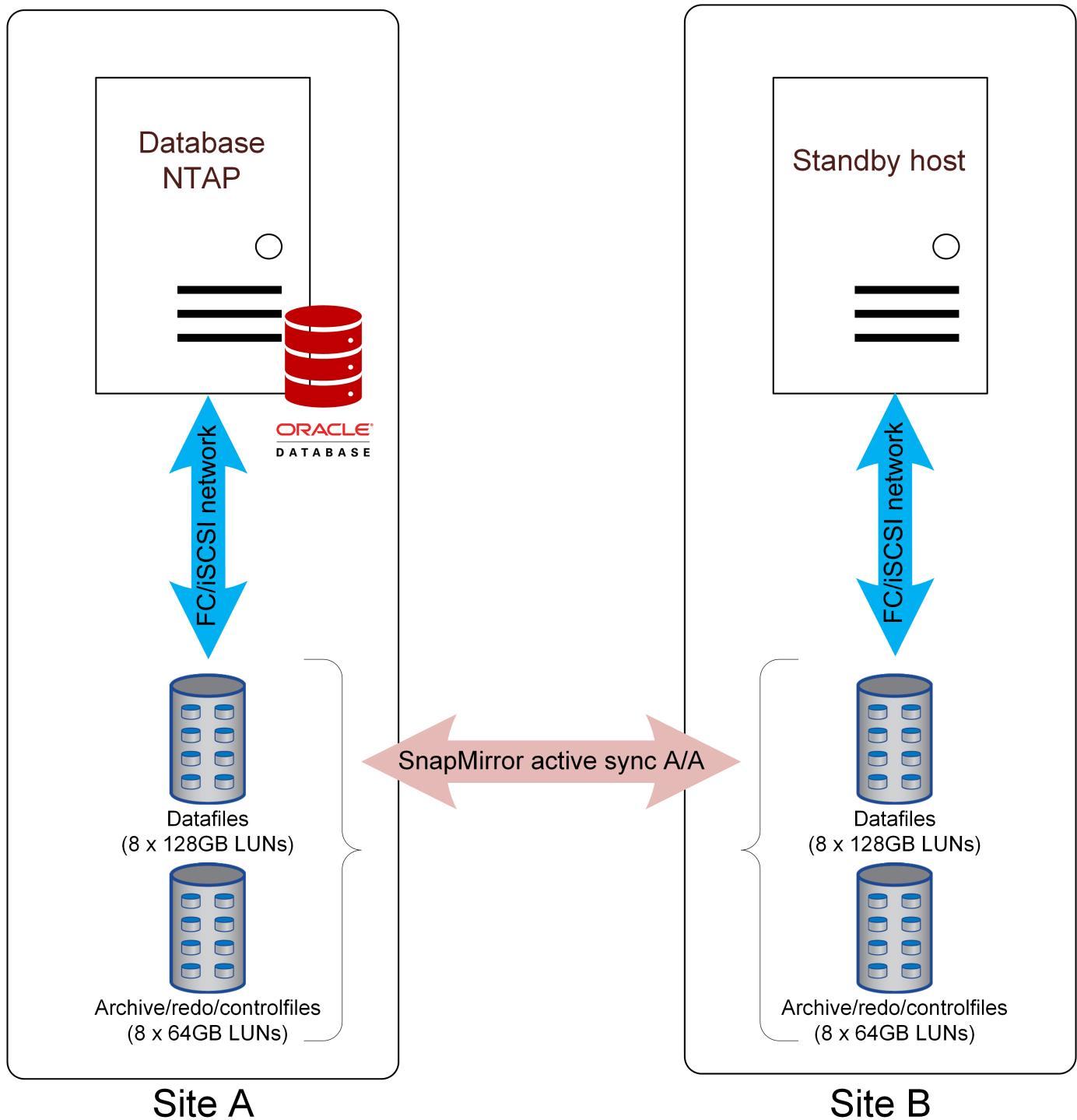
Si le RTO était plus strict, l'automatisation actif-passif de base via des scripts ou des clusters comme Pacemaker ou Ansible améliorerait le délai de basculement. Par exemple, VMware HA peut être configuré pour détecter une panne de VM sur le site principal et activer cette dernière sur le site distant.

Enfin, pour un basculement extrêmement rapide, Oracle RAC peut être déployé sur plusieurs sites. L'objectif de délai de restauration serait essentiellement égal à zéro, car la base de données serait en ligne et disponible à tout moment sur les deux sites.

Instance unique Oracle

Les exemples décrits ci-dessous illustrent certaines des nombreuses options de déploiement des bases de données Oracle Single instance avec la réplication SnapMirror

Active Sync.



Basculement avec un système d'exploitation préconfiguré

La synchronisation active SnapMirror fournit une copie synchrone des données au niveau du site de reprise d'activité. Toutefois, la mise à disposition des données requiert un système d'exploitation et les applications associées. L'automatisation de base peut considérablement améliorer le délai de basculement de l'environnement global. Les produits Clusterware tels que Pacemaker sont souvent utilisés pour créer un cluster sur les sites et, dans la plupart des cas, le processus de basculement peut être piloté par des scripts simples.

En cas de perte des nœuds principaux, le cluster (ou les scripts) mettra les bases de données en ligne sur le site secondaire. Une option consiste à créer des serveurs de secours préconfigurés pour les ressources SAN qui constituent la base de données. En cas de défaillance du site principal, le logiciel de mise en cluster ou l'alternative scriptée effectue une séquence d'actions similaires à celles décrites ci-dessous :

1. Détection d'une défaillance du site principal
2. Effectuez la détection des LUN FC ou iSCSI
3. Montage de systèmes de fichiers et/ou montage de groupes de disques ASM
4. Démarrage de la base de données

Cette approche doit avant tout se passer d'un système d'exploitation en cours d'exécution sur le site distant. Elles doivent être préconfigurées avec des binaires Oracle, ce qui signifie également que des tâches telles que l'application de correctifs Oracle doivent être effectuées sur les sites principal et de secours. Les binaires Oracle peuvent également être mis en miroir vers le site distant et montés en cas d'incident.

La procédure d'activation réelle est simple. Les commandes telles que la découverte de LUN ne nécessitent que quelques commandes par port FC. Le montage du système de fichiers n'est rien de plus qu'une `mount` commande et les bases de données et ASM peuvent être démarrés et arrêtés sur l'interface de ligne de commande à l'aide d'une seule commande.

Basculement avec un système d'exploitation virtualisé

Le basculement des environnements de base de données peut être étendu pour inclure le système d'exploitation lui-même. En théorie, ce basculement peut être effectué avec des LUN de démarrage, mais le plus souvent avec un système d'exploitation virtualisé. La procédure est similaire aux étapes suivantes :

1. Détection d'une défaillance du site principal
2. Montage des datastores hébergeant les machines virtuelles du serveur de base de données
3. Démarrage des machines virtuelles
4. Démarrage manuel des bases de données ou configuration des machines virtuelles pour démarrer automatiquement les bases de données.

Par exemple, un cluster ESX peut couvrir des sites. En cas d'incident, les machines virtuelles peuvent être mises en ligne sur le site de reprise après incident après le basculement.

Protection contre les défaillances du stockage

Le diagramme ci-dessus montre l'utilisation de "[accès non uniforme](#)", où le SAN n'est pas étendu entre les sites. Cela peut être plus simple à configurer et, dans certains cas, peut être la seule option étant donné les fonctionnalités SAN actuelles, mais cela signifie également que la défaillance du système de stockage principal entraînerait une panne de la base de données jusqu'à ce que l'application ait été ratée.

Pour une résilience supplémentaire, la solution pourrait être déployée avec "[accès uniforme](#)". Cela permettrait aux applications de continuer à fonctionner en utilisant les chemins annoncés à partir du site opposé.

RAC étendu Oracle

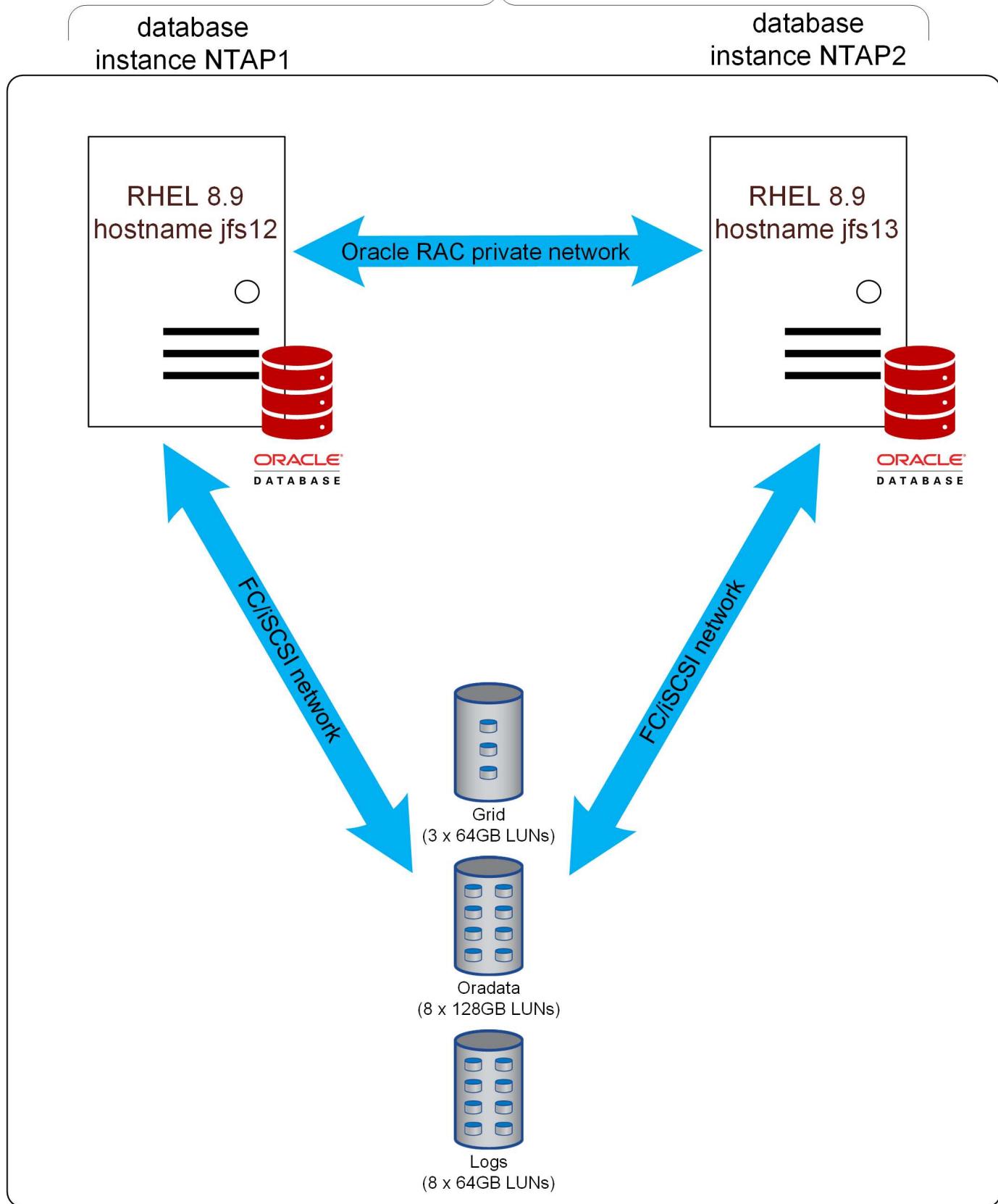
De nombreux clients optimisent leur RTO en étendant un cluster Oracle RAC sur plusieurs sites, offrant une configuration entièrement active/active. La conception globale devient plus complexe car elle doit inclure la gestion du quorum d'Oracle RAC.

La mise en cluster RAC étendue traditionnelle s'est appuyée sur la mise en miroir ASM pour assurer la protection des données. Cette approche fonctionne, mais elle implique également de nombreuses étapes manuelles de configuration et entraîne une surcharge de l'infrastructure réseau. À l'inverse, la réplication des données peut être prise en charge par la synchronisation active SnapMirror, ce qui simplifie considérablement la solution. Les opérations telles que la synchronisation, la resynchronisation après les interruptions, les basculements et la gestion du quorum sont plus simples. En outre, le SAN n'a pas besoin d'être distribué entre les sites, ce qui simplifie la conception et la gestion du SAN.

La réplication

Pour comprendre la fonctionnalité RAC sur SnapMirror Active Sync, il est essentiel de considérer le stockage comme un ensemble unique de LUN hébergés sur un stockage en miroir. Par exemple :

Database NTAP



Il n'y a pas de copie principale ou miroir. Pour schématiser, il n'y a qu'une seule copie de chaque LUN et cette LUN est disponible sur les chemins SAN situés sur deux systèmes de stockage différents. Du point de vue de l'hôte, il n'y a pas de basculement de stockage ; il y a des changements de chemin. Plusieurs défaillances

peuvent entraîner la perte de certains chemins vers la LUN, tandis que les autres chemins restent en ligne. La synchronisation active SnapMirror garantit la disponibilité des mêmes données sur tous les chemins opérationnels.

Configuration de stockage sous-jacente

Dans cet exemple de configuration, les disques ASM sont configurés de la même manière que dans n'importe quelle configuration RAC à site unique sur le stockage d'entreprise. Étant donné que le système de stockage assure la protection des données, la redondance ASM externe est utilisée.

Accès uniforme ou non informé

L'élément le plus important à prendre en compte avec Oracle RAC sur SnapMirror Active Sync est de savoir s'il faut utiliser un accès uniforme ou non.

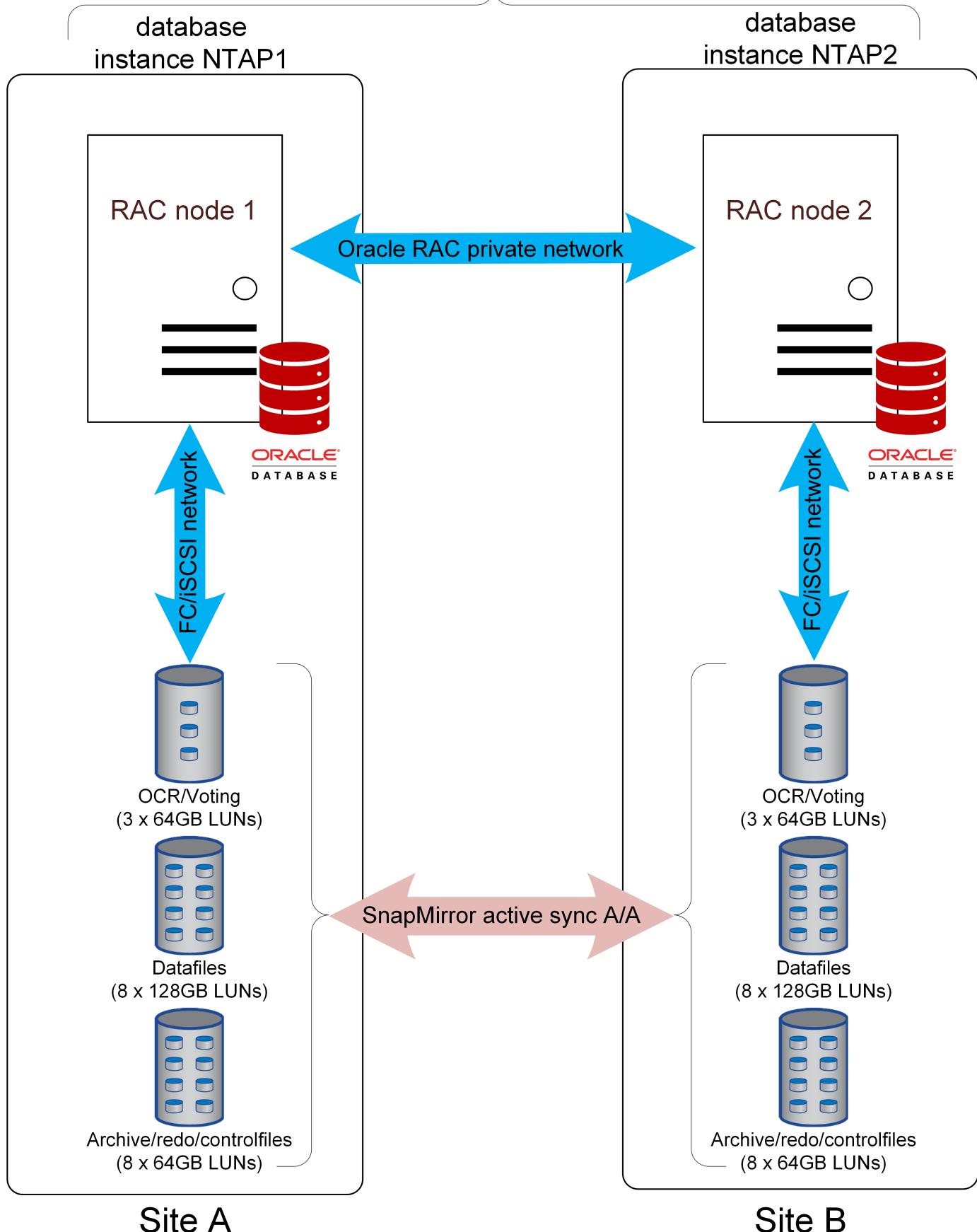
Un accès uniforme signifie que chaque hôte peut voir les chemins sur les deux clusters. L'accès non uniforme signifie que les hôtes peuvent uniquement voir les chemins vers le cluster local.

Aucune de ces options n'est spécifiquement recommandée ou déconseillée. Certains clients ont facilement accès à la fibre noire pour connecter les sites, d'autres ne disposent pas d'une telle connectivité ou leur infrastructure SAN ne prend pas en charge l'ISL longue distance.

Accès non uniforme

L'accès non uniforme est plus simple à configurer du point de vue du SAN.

Database NTAP



L'inconvénient principal de cette "[accès non uniforme](#)" approche est que la perte de la connectivité ONTAP site à site ou la perte d'un système de stockage entraînera la perte des instances de base de données sur un site. Cela n'est évidemment pas souhaitable, mais cela peut constituer un risque acceptable en échange d'une configuration SAN plus simple.

Accès uniforme

L'accès uniforme requiert l'extension du SAN sur les sites. Le principal avantage est que la perte d'un système de stockage n'entraîne pas la perte d'une instance de base de données. Au lieu de cela, cela entraînerait une modification des chemins d'accès multiples dans lesquels les chemins sont actuellement utilisés.

Il existe plusieurs façons de configurer l'accès non uniforme.

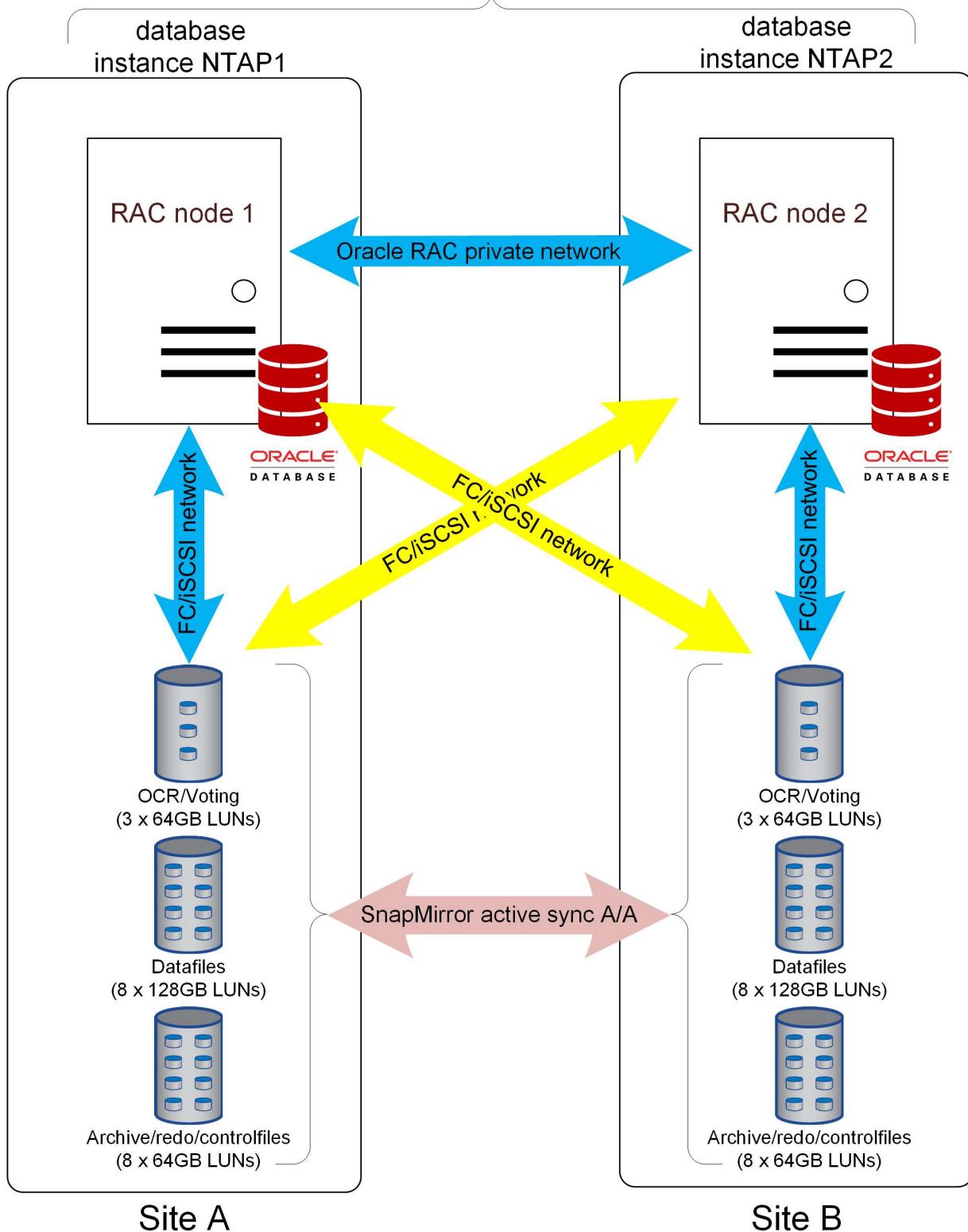


Dans les schémas ci-dessous, il existe également des chemins actifs mais non optimisés qui seraient utilisés en cas de défaillances simples du contrôleur, mais ces chemins ne sont pas affichés dans l'intérêt de simplifier les diagrammes.

AFF avec paramètres de proximité

En cas de latence importante entre les sites, les systèmes AFF peuvent être configurés avec des paramètres de proximité des hôtes. Cela permet à chaque système de stockage d'identifier les hôtes locaux et distants, et d'attribuer les priorités de chemin en conséquence.

Database NTAP



Active/Optimized Path

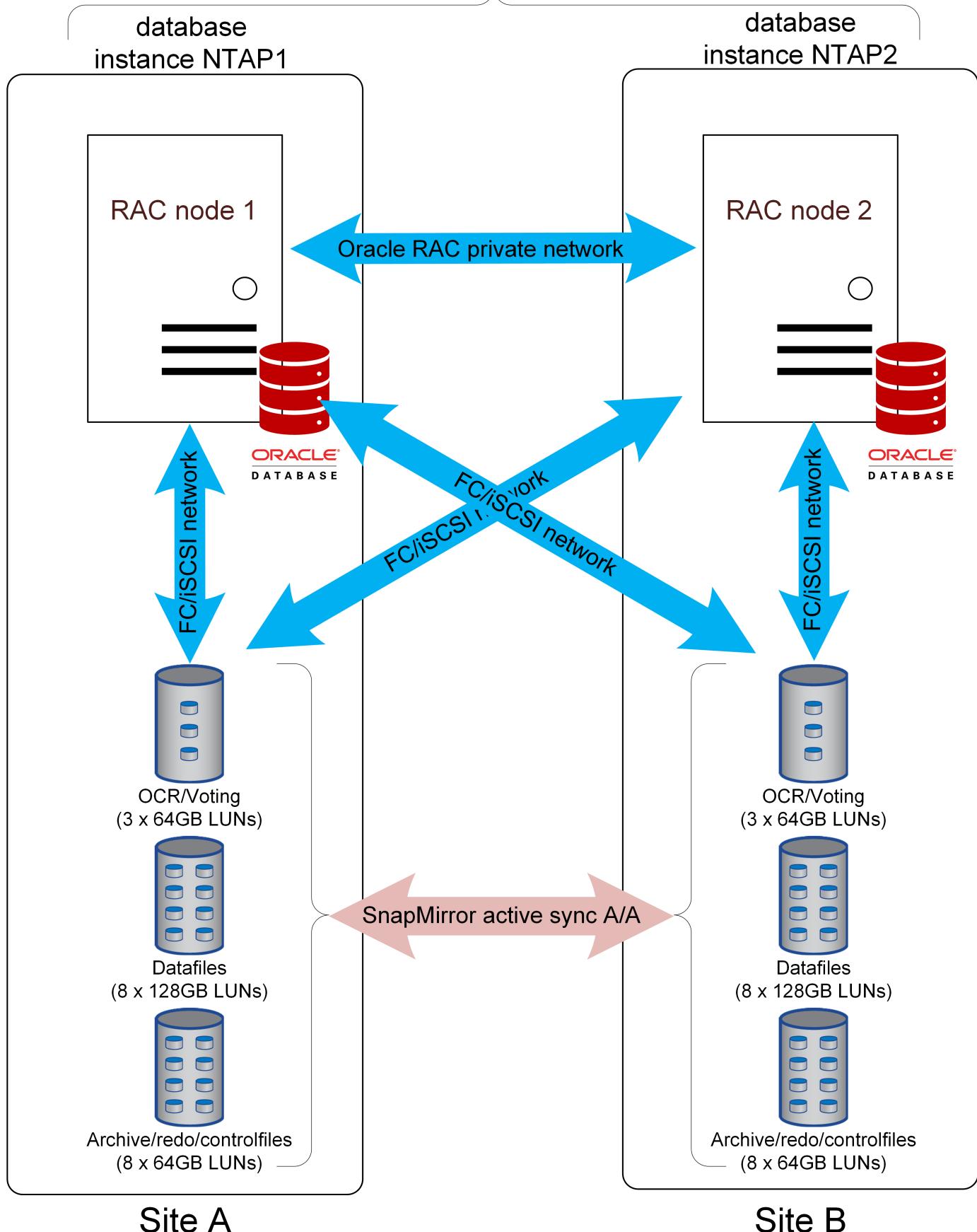
Active Path

En fonctionnement normal, chaque instance Oracle utilisera de préférence les chemins locaux actifs/optimisés. Par conséquent, toutes les lectures seront traitées par la copie locale des blocs. La latence est ainsi la plus faible possible. Les E/S d'écriture sont envoyées de la même manière vers le contrôleur local. L'E/S doit toujours être répliquée avant d'être reconnue, ce qui entraîne toujours une latence supplémentaire en traversant le réseau site à site, mais cela ne peut pas être évité dans une solution de réplication synchrone.

ASA / AFF sans paramètres de proximité

S'il n'y a pas de latence significative entre les sites, les systèmes AFF peuvent être configurés sans paramètres de proximité des hôtes, ou ASA peut être utilisé.

Database NTAP



Chaque hôte pourra utiliser tous les chemins opérationnels sur les deux systèmes de stockage. Cela améliore considérablement les performances en permettant à chaque hôte d'exploiter le potentiel de performance de deux clusters, et non d'un seul.

Avec ASA, non seulement tous les chemins vers les deux clusters sont considérés comme actifs et optimisés, mais les chemins sur les contrôleurs partenaires sont également actifs. Il en résulte des chemins SAN entièrement actifs sur l'ensemble du cluster, à tout moment.



Les systèmes ASA peuvent également être utilisés dans une configuration d'accès non uniforme. Étant donné qu'il n'existe aucun chemin entre les sites, les performances ne seraient pas améliorées par le franchissement de l'ISL par les E/S.

Disjoncteur d'attache RAC

Bien que le RAC étendu utilisant la synchronisation active SnapMirror soit une architecture symétrique par rapport aux E/S, il existe une exception qui est connectée à la gestion du split-brain.

Que se passe-t-il si le lien de réplication est perdu et qu'aucun des sites n'a le quorum ? Que doit-on faire ? Cette question s'applique à la fois au comportement d'Oracle RAC et de ONTAP. Si les modifications ne peuvent pas être répliquées sur tous les sites et que vous souhaitez reprendre les opérations, l'un des sites devra survivre et l'autre site devra être indisponible.

Le système "[Médiateur de ONTAP](#)" répond à cette exigence au niveau de la couche ONTAP. Il existe plusieurs options pour le tricover RAC.

Disjoncteurs Oracle

La meilleure méthode pour gérer les risques Oracle RAC split-brain consiste à utiliser un nombre impair de nœuds RAC, de préférence à l'aide d'un Tiebreaker 3rd site. Si un troisième site n'est pas disponible, l'instance Tiebreaker pourrait être placée sur un site des deux sites, ce qui la désignerait en fait un site de survivant préféré.

Oracle et CSS_Critical

Avec un nombre pair de nœuds, le comportement par défaut d'Oracle RAC est que l'un des nœuds du cluster sera considéré plus important que les autres nœuds. Le site avec ce noeud de priorité supérieure survivra à l'isolation du site tandis que les nœuds de l'autre site seront supprimés. La hiérarchisation est basée sur plusieurs facteurs, mais vous pouvez également contrôler ce comportement à l'aide du `css_critical` paramètre.

Dans l'"[exemple](#)" architecture, les noms d'hôte des nœuds RAC sont jfs12 et jfs13. Les paramètres actuels de `css_critical` sont les suivants :

```
[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.
```

Si vous voulez que le site avec jfs12 soit le site préféré, définissez cette valeur sur oui sur un site Un noeud et

redémarrez les services.

```
[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.
```

Scénarios d'échec

Présentation

La planification d'une architecture complète d'applications de synchronisation active SnapMirror nécessite de comprendre comment les SM-AS répondront dans divers scénarios de basculement planifiés et non planifiés.

Pour les exemples suivants, supposons que le site A est configuré comme le site préféré.

Perte de la connectivité de réPLICATION

Si la réPLICATION SM-AS est interrompue, l'E/S d'écriture ne peut pas être terminée, car un cluster ne peut pas répliquer les modifications sur le site opposé.

Site A (site préféré)

Le résultat de l'échec de la liaison de réPLICATION sur le site préféré sera une pause d'environ 15 secondes dans le traitement des E/S d'écriture, car ONTAP relance les opérations d'écriture répliquées avant de déterminer que la liaison de réPLICATION est véritablement inaccessible. Au bout de 15 secondes, le site A du système reprend le traitement des E/S de lecture et d'écriture. Les chemins SAN ne changent pas et les LUN restent en ligne.

Site B

Le site B n'étant pas le site privilégié de synchronisation active SnapMirror, ses chemins de LUN deviennent indisponibles au bout de 15 secondes environ.

Panne du système de stockage

Le résultat d'une défaillance du système de stockage est presque identique au résultat de la perte du lien de réplication. Le site survivant devrait subir une pause d'E/S d'environ 15 seconde. Une fois cette période de 15 secondes écoulée, l'E/S reprend sur ce site comme d'habitude.

Perte du médiateur

Le service médiateur ne contrôle pas directement les opérations de stockage. Il fonctionne comme un chemin de contrôle alternatif entre les clusters. Il existe principalement pour automatiser le basculement sans les risques associés à un scénario « split-brain ». En conditions normales de fonctionnement, chaque cluster réplique les modifications apportées à son partenaire et chaque cluster peut donc vérifier que le cluster partenaire est en ligne et qu'il transmet les données. Si le lien de réplication échoue, la réplication s'arrête.

La raison pour laquelle un médiateur est nécessaire pour un basculement automatisé sécurisé est parce qu'il serait autrement impossible à un cluster de stockage de déterminer si la perte de la communication bidirectionnelle était le résultat d'une panne du réseau ou d'une défaillance réelle du stockage.

Le médiateur fournit un chemin alternatif pour chaque cluster afin de vérifier l'état de santé de son partenaire. Les scénarios sont les suivants :

- Si un cluster peut contacter directement son partenaire, les services de réplication sont opérationnels. Aucune action requise.
- Si un site privilégié ne peut pas contacter son partenaire directement ou via le médiateur, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site préféré va ensuite publier l'état RPO=0 et continuer à traiter les E/S en lecture et en écriture.
- Si un site non préféré ne peut pas contacter directement son partenaire, mais peut le contacter via le médiateur, il mettra ses chemins hors ligne et attend le retour de la connexion de réplication.
- Si un site non privilégié ne peut pas contacter son partenaire directement ou via un médiateur opérationnel, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site non privilégié va ensuite publier l'état RPO=0 et continuer le traitement des E/S en lecture et en écriture. Il assumera le rôle de la source de réplication et deviendra le nouveau site préféré.

Si le médiateur n'est pas disponible :

- En cas de défaillance des services de réplication, quelle qu'en soit la raison, y compris la défaillance du site ou du système de stockage non privilégié, le site préféré libère l'état RPO=0 et reprend le traitement des E/S de lecture et d'écriture. Le site non préféré mettra ses chemins hors ligne.
- La défaillance du site préféré entraînera une panne, car le site non préféré ne pourra pas vérifier que le site opposé est réellement hors ligne et, par conséquent, il ne serait pas sûr que le site non préféré puisse reprendre ses services.

Restauration des services

Après résolution d'une panne, par exemple lors de la restauration de la connectivité site à site ou de la mise sous tension d'un système défaillant, les terminaux de synchronisation active SnapMirror détectent automatiquement la présence d'une relation de réplication défectueuse et la raverront à l'état RPO=0. Une fois la réplication synchrone rétablie, les chemins défaillants se reconnectent.

Dans de nombreux cas, les applications en cluster détectent automatiquement le retour des chemins défaillants, et ces applications sont également reconnectées. Dans d'autres cas, une analyse SAN au niveau de l'hôte peut être nécessaire ou les applications doivent être reconnectées manuellement. Cela dépend de l'application et de la façon dont elle est configurée et, en général, de telles tâches peuvent être facilement automatisées. La fonctionnalité ONTAP elle-même est dotée d'une fonctionnalité d'autorétablissement et ne nécessite aucune intervention de l'utilisateur pour reprendre les opérations de stockage avec un objectif de point de récupération de 0.

Basculement manuel

La modification du site préféré nécessite une opération simple. L'E/S s'interrompt pendant une ou deux secondes car l'autorité sur le comportement de réplication change entre les clusters, mais l'E/S n'est pas affectée.

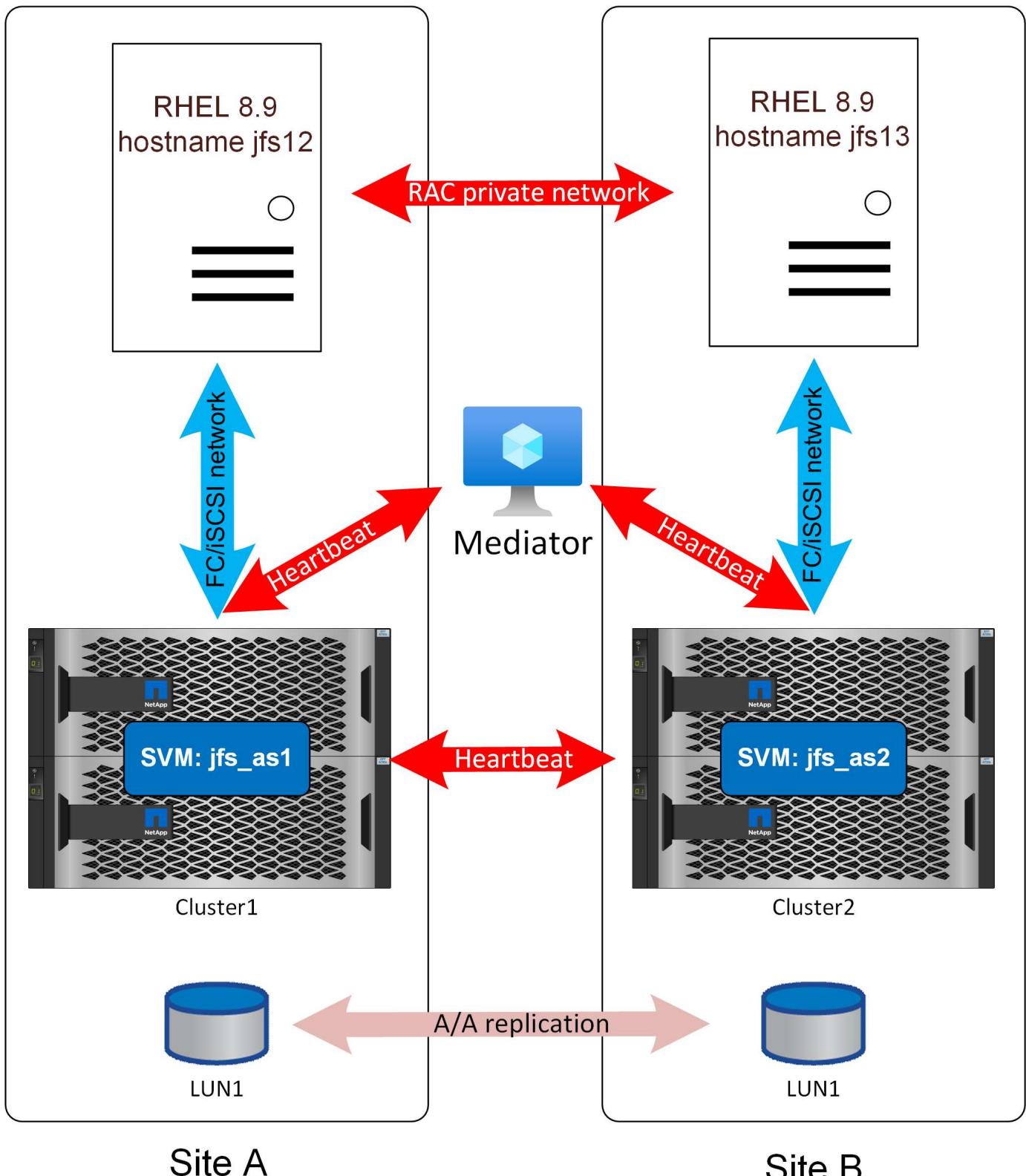
Exemple d'architecture

Les exemples détaillés de défaillances présentés dans cette section sont basés sur l'architecture présentée ci-dessous.



Il ne s'agit que de l'une des nombreuses options pour les bases de données Oracle sur la synchronisation active SnapMirror. Cette conception a été choisie parce qu'elle illustre certains des scénarios les plus complexes.

Dans cette conception, supposons que le site A est défini sur "["site préféré"](#)".



Échec de l'interconnexion du RAC

La perte du lien de réplication RAC Oracle produira un résultat similaire à la perte de la connectivité SnapMirror, sauf que les délais d'expiration seront plus courts par défaut. Dans les paramètres par défaut, un nœud RAC Oracle attend 200 secondes après une

perte de connectivité du stockage avant d'être supprimé, mais il n'attend que 30 secondes après la perte du signal de détection du réseau RAC.

Les messages CRS sont similaires à ceux indiqués ci-dessous. Vous pouvez voir le délai d'expiration de 30 secondes. Comme css_Critical a été défini sur jfs12, situé sur le site A, ce sera le site pour survivre et jfs13 sur le site B sera supprimé.

```
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'.
```

Échec de communication SnapMirror

Si la liaison de réplication SnapMirror active Sync, l'E/S d'écriture ne peut pas être terminée, car un cluster ne peut pas répliquer les modifications sur le site opposé.

Site A

Le site A qui présente une défaillance de liaison de réplication entraînera une pause d'environ 15 secondes dans le traitement des E/S d'écriture au fur et à mesure que ONTAP tente de répliquer des écritures avant de déterminer que la liaison de réplication est réellement inutilisable. Au bout de 15 secondes, le cluster ONTAP sur le site A reprend le traitement des E/S de lecture et d'écriture. Les chemins SAN ne changent pas et les LUN restent en ligne.

Site B

Le site B n'étant pas le site privilégié de synchronisation active SnapMirror, ses chemins de LUN deviennent indisponibles au bout de 15 secondes environ.

Le lien de réplication a été coupé à l'horodatage 15:19:44. Le premier avertissement d'Oracle RAC arrive 100 secondes plus tard lorsque le délai d'expiration de 200 secondes (contrôlé par le paramètre Oracle RAC `disktimeout`) approche.

```
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 CRSD
resources.
```

Une fois que le délai d'expiration du disque de vote de 200 secondes a été atteint, ce nœud RAC Oracle s'expulse automatiquement du cluster et redémarre.

Échec total de l'interconnectivité réseau

Si la liaison de réplication entre les sites est totalement perdue, la synchronisation active SnapMirror et la connectivité RAC Oracle seront interrompues.

La détection d'Oracle RAC à cerveau divisé dépend du pulsation du stockage Oracle RAC. Si la perte de la connectivité site à site entraîne la perte simultanée du signal de détection du réseau RAC et des services de réplication du stockage, les sites RAC ne pourront pas communiquer entre sites via l'interconnexion RAC ou les disques de vote RAC. Le résultat d'un ensemble de nœuds à numéro pair peut être l'exclusion des deux sites sous les paramètres par défaut. Le comportement exact dépend de la séquence des événements et de la synchronisation des sondages de pulsation du réseau RAC et du disque.

Le risque d'une panne sur deux sites peut être résolu de deux manières. Tout d'abord, une "[disjoncteur d'attache](#)" configuration peut être utilisée.

Si aucun site tiers n'est disponible, ce risque peut être résolu en ajustant le paramètre `misscount` sur le cluster RAC. Sous les valeurs par défaut, le délai d'expiration de la pulsation réseau du RAC est de 30 secondes. Il est généralement utilisé par RAC pour identifier les nœuds RAC défaillants et les supprimer du cluster. Il dispose également d'une connexion à la pulsation du disque de vote.

Si, par exemple, le conduit transportant le trafic intersite pour Oracle RAC et les services de réplication de stockage est coupé par une pelle rétro, le compte à rebours des erreurs de 30 secondes commence. Si le nœud du site RAC préféré ne peut pas rétablir le contact avec le site opposé dans les 30 secondes et qu'il ne peut pas utiliser les disques de vote pour confirmer que le site opposé est en panne dans la même fenêtre de 30 secondes, les nœuds du site préféré seront également supprimés. Il en résulte une interruption complète de la base de données.

Selon le moment où l'interrogation du compte erroné se produit, 30 secondes peuvent ne pas suffire à la temporisation de la synchronisation active SnapMirror et à permettre au stockage du site préféré de reprendre les services avant l'expiration de la fenêtre de 30 secondes. Cette fenêtre de 30 secondes peut être augmentée.

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

Cette valeur permet au système de stockage sur le site préféré de reprendre les opérations avant que le délai d'erreur n'expire. Le résultat sera alors la suppression uniquement des nœuds sur le site où les chemins de LUN ont été supprimés. Exemple ci-dessous :

```
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
```

Le support Oracle déconseille fortement de modifier les paramètres misscount ou disktimeout pour résoudre les problèmes de configuration. Toutefois, la modification de ces paramètres peut s'avérer justifiée et inévitable dans de nombreux cas, notamment dans les configurations de démarrage SAN, de virtualisation et de réplication du stockage. Si, par exemple, vous avez rencontré des problèmes de stabilité avec un réseau SAN

ou IP qui ont entraîné des expulsions RAC, vous devez résoudre le problème sous-jacent et ne pas facturer les valeurs de l'erreur de décompte ou du dépassement de disque. La modification des délais pour résoudre les erreurs de configuration masque un problème et non pas résout un problème. La modification de ces paramètres pour configurer correctement un environnement RAC basé sur les aspects de conception de l'infrastructure sous-jacente est différente et est conforme aux instructions de support Oracle. Avec le démarrage SAN, il est courant d'ajuster misscount jusqu'à 200 pour correspondre au disktimeout. Voir "[ce lien](#)" pour plus d'informations.

Panne du site

Le résultat d'une défaillance du site ou du système de stockage est presque identique au résultat de la perte du lien de réPLICATION. Le site survivant doit subir une pause d'E/S d'environ 15 secondes sur les écritures. Une fois cette période de 15 secondes écoulée, l'E/S reprend sur ce site comme d'habitude.

Si seul le système de stockage a été affecté, le nœud Oracle RAC sur le site en panne perdra les services de stockage et entrera le même compte à rebours de 200 secondes avant la suppression et le redémarrage suivant.

```

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 CRSD
resources.

```

L'état du chemin SAN sur le nœud RAC qui a perdu des services de stockage se présente comme suit :

```

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

```

L'hôte linux a détecté la perte des chemins beaucoup plus rapidement que 200 secondes, mais du point de vue de la base de données, les connexions client à l'hôte sur le site défaillant seront toujours bloquées pendant 200 secondes sous les paramètres Oracle RAC par défaut. Les opérations complètes de la base de données ne reprennent qu'une fois la suppression terminée.

Pendant ce temps, le nœud Oracle RAC sur le site opposé enregistre la perte de l'autre nœud RAC. Dans le cas contraire, le système continue de fonctionner normalement.

```
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
```

Défaillance du médiateur

Le service médiateur ne contrôle pas directement les opérations de stockage. Il fonctionne comme un chemin de contrôle alternatif entre les clusters. Il existe principalement pour automatiser le basculement sans les risques associés à un scénario « split-brain ».

En conditions normales de fonctionnement, chaque cluster réplique les modifications apportées à son partenaire et chaque cluster peut donc vérifier que le cluster partenaire est en ligne et qu'il transmet les données. Si le lien de réplication échoue, la réplication s'arrête.

Un médiateur est nécessaire pour des opérations automatisées sécurisées, car il serait autrement impossible pour les clusters de stockage de déterminer si la perte de la communication bidirectionnelle était due à une panne du réseau ou à une défaillance réelle du stockage.

Le médiateur fournit un chemin alternatif pour chaque cluster afin de vérifier l'état de santé de son partenaire. Les scénarios sont les suivants :

- Si un cluster peut contacter directement son partenaire, les services de réplication sont opérationnels. Aucune action requise.
- Si un site privilégié ne peut pas contacter son partenaire directement ou via le médiateur, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site préféré va ensuite publier l'état RPO=0 et continuer à traiter les E/S en lecture et en écriture.
- Si un site non préféré ne peut pas contacter directement son partenaire, mais peut le contacter via le médiateur, il mettra ses chemins hors ligne et attend le retour de la connexion de réplication.

- Si un site non privilégié ne peut pas contacter son partenaire directement ou via un médiateur opérationnel, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site non privilégié va ensuite publier l'état RPO=0 et continuer le traitement des E/S en lecture et en écriture. Il assumera le rôle de la source de réplication et deviendra le nouveau site préféré.

Si le médiateur n'est pas disponible :

- En cas de défaillance des services de réplication, quelle qu'en soit la raison, le site préféré libère l'état RPO=0 et reprend le traitement des E/S en lecture et en écriture. Le site non préféré mettra ses chemins hors ligne.
- La défaillance du site préféré entraînera une panne, car le site non préféré ne pourra pas vérifier que le site opposé est réellement hors ligne et, par conséquent, il ne serait pas sûr que le site non préféré puisse reprendre ses services.

Restauration du service

SnapMirror propose une fonctionnalité d'autorétablissement. La synchronisation active SnapMirror détecte automatiquement la présence d'une relation de réplication défectueuse et la ramène à un état RPO=0. Une fois la réplication synchrone rétablie, les chemins reviennent en ligne.

Dans de nombreux cas, les applications en cluster détectent automatiquement le retour des chemins défaillants, et ces applications sont également reconnectées. Dans d'autres cas, une analyse SAN au niveau de l'hôte peut être nécessaire ou les applications doivent être reconnectées manuellement.

Cela dépend de l'application et de la façon dont elle est configurée et, en général, ces tâches peuvent être facilement automatisées. La synchronisation active SnapMirror elle-même est auto-fixing et ne nécessite aucune intervention de l'utilisateur pour reprendre les opérations de stockage avec un objectif de point de récupération de 0 une fois l'alimentation et la connectivité restaurées.

Basculement manuel

Le terme « basculement » ne fait pas référence au sens de la réplication avec la synchronisation active SnapMirror, car il s'agit d'une technologie de réplication bidirectionnelle. En revanche, le terme « basculement » désigne le système de stockage qui sera le site privilégié en cas de défaillance.

Par exemple, vous pouvez effectuer un basculement pour modifier le site préféré avant d'arrêter un site pour des raisons de maintenance ou avant d'effectuer un test de reprise après incident.

La modification du site préféré nécessite une opération simple. L'E/S s'interrompt pendant une ou deux secondes car l'autorité sur le comportement de réplication change entre les clusters, mais l'E/S n'est pas affectée.

Exemple d'interface graphique :

Relationships

Local destinations

Local sources

Search

Download

Show/hide

Filter

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

Exemple de modification via l'interface de ligne de commande :

```
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	Destination		Error			
Path	Path	Type	Status	start-time	end-time	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
```

Informations sur le copyright

Copyright © 2026 NetApp, Inc. Tous droits réservés. Imprimé aux États-Unis. Aucune partie de ce document protégé par copyright ne peut être reproduite sous quelque forme que ce soit ou selon quelque méthode que ce soit (graphique, électronique ou mécanique, notamment par photocopie, enregistrement ou stockage dans un système de récupération électronique) sans l'autorisation écrite préalable du détenteur du droit de copyright.

Les logiciels dérivés des éléments NetApp protégés par copyright sont soumis à la licence et à l'avis de non-responsabilité suivants :

CE LOGICIEL EST FOURNI PAR NETAPP « EN L'ÉTAT » ET SANS GARANTIES EXPRESSES OU TACITES, Y COMPRIS LES GARANTIES TACITES DE QUALITÉ MARCHANDE ET D'ADÉQUATION À UN USAGE PARTICULIER, QUI SONT EXCLUES PAR LES PRÉSENTES. EN AUCUN CAS NETAPP NE SERA TENU POUR RESPONSABLE DE DOMMAGES DIRECTS, INDIRECTS, ACCESSOIRES, PARTICULIERS OU EXEMPLAIRES (Y COMPRIS L'ACHAT DE BIENS ET DE SERVICES DE SUBSTITUTION, LA PERTE DE JOUSSANCE, DE DONNÉES OU DE PROFITS, OU L'INTERRUPTION D'ACTIVITÉ), QUELLES QU'EN SOIENT LA CAUSE ET LA DOCTRINE DE RESPONSABILITÉ, QU'IL S'AGISSE DE RESPONSABILITÉ CONTRACTUELLE, STRICTE OU DÉLICTUELLE (Y COMPRIS LA NÉGLIGENCE OU AUTRE) DÉCOULANT DE L'UTILISATION DE CE LOGICIEL, MÊME SI LA SOCIÉTÉ A ÉTÉ INFORMÉE DE LA POSSIBILITÉ DE TELS DOMMAGES.

NetApp se réserve le droit de modifier les produits décrits dans le présent document à tout moment et sans préavis. NetApp décline toute responsabilité découlant de l'utilisation des produits décrits dans le présent document, sauf accord explicite écrit de NetApp. L'utilisation ou l'achat de ce produit ne concède pas de licence dans le cadre de droits de brevet, de droits de marque commerciale ou de tout autre droit de propriété intellectuelle de NetApp.

Le produit décrit dans ce manuel peut être protégé par un ou plusieurs brevets américains, étrangers ou par une demande en attente.

LÉGENDE DE RESTRICTION DES DROITS : L'utilisation, la duplication ou la divulgation par le gouvernement sont sujettes aux restrictions énoncées dans le sous-paragraphe (b)(3) de la clause Rights in Technical Data-Noncommercial Items du DFARS 252.227-7013 (février 2014) et du FAR 52.227-19 (décembre 2007).

Les données contenues dans les présentes se rapportent à un produit et/ou service commercial (tel que défini par la clause FAR 2.101). Il s'agit de données propriétaires de NetApp, Inc. Toutes les données techniques et tous les logiciels fournis par NetApp en vertu du présent Accord sont à caractère commercial et ont été exclusivement développés à l'aide de fonds privés. Le gouvernement des États-Unis dispose d'une licence limitée irrévocable, non exclusive, non cessible, non transférable et mondiale. Cette licence lui permet d'utiliser uniquement les données relatives au contrat du gouvernement des États-Unis d'après lequel les données lui ont été fournies ou celles qui sont nécessaires à son exécution. Sauf dispositions contraires énoncées dans les présentes, l'utilisation, la divulgation, la reproduction, la modification, l'exécution, l'affichage des données sont interdits sans avoir obtenu le consentement écrit préalable de NetApp, Inc. Les droits de licences du Département de la Défense du gouvernement des États-Unis se limitent aux droits identifiés par la clause 252.227-7015(b) du DFARS (février 2014).

Informations sur les marques commerciales

NETAPP, le logo NETAPP et les marques citées sur le site <http://www.netapp.com/TM> sont des marques déposées ou des marques commerciales de NetApp, Inc. Les autres noms de marques et de produits sont des marques commerciales de leurs propriétaires respectifs.