

Résilience de la charge de travail avec EMC VPLEX

Meilleures pratiques de planification

Résumé

Ce livre blanc fournit une brève introduction à EMC[®] VPLEX[™] et décrit comment assurer au datacenter une résilience accrue de la charge de travail grâce à VPLEX. Il propose des recommandations quant aux meilleures pratiques pour réaliser des déploiements à haute disponibilité et explique comment VPLEX gère divers scénarios de panne.

Copyright © 2010 EMC Corporation. Tous droits réservés.

EMC estime que les informations figurant dans ce document sont exactes à la date de publication. Ces informations sont modifiables sans préavis.

LES INFORMATIONS CONTENUES DANS CETTE PUBLICATION SONT FOURNIES « EN L'ÉTAT ». EMC CORPORATION NE FOURNIT AUCUNE DÉCLARATION OU GARANTIE D'AUCUNE SORTE CONCERNANT LES INFORMATIONS CONTENUES DANS CETTE PUBLICATION ET REJETTE PLUS SPÉCIALEMENT TOUTE GARANTIE IMPLICITE DE QUALITÉ COMMERCIALE OU D'ADÉQUATION À UNE UTILISATION PARTICULIÈRE.

L'utilisation, la copie et la diffusion de tout logiciel EMC décrit dans cette publication nécessitent une licence logicielle en cours de validité.

Pour obtenir la liste actualisée des noms de produits, consultez la rubrique des marques EMC via le lien Législation, sur emc2.fr.

Toutes les autres marques citées dans le présent document sont la propriété de leurs détenteurs respectifs.

Référence h7138.1

Table des matières

Résumé analytique	4
Introduction	4
Public concerné	4
Présentation de la technologie VPLEX	4
L'architecture en cluster d'EMC VPLEX	6
Virtualisation des périphériques avec VPLEX.....	7
Présentation du matériel VPLEX	7
Présentation du déploiement.....	9
Déploiement VPLEX Local.....	9
Cas d'utilisation d'un déploiement VPLEX Local	10
Déploiement de VPLEX Metro au sein d'un datacenter	10
Cas d'utilisation d'un déploiement VPLEX Metro au sein d'un datacenter	11
Déploiement VPLEX Metro entre des datacenters	11
Cas d'utilisation d'un déploiement VPLEX Metro entre des datacenters.....	12
Résilience de la charge de travail.....	12
Arrêts des baies de stockage.....	13
Meilleures pratiques	13
Pannes du réseau SAN	14
Bonne pratique	14
Défaillance des composants VPLEX	15
Défaillance de port Fibre Channel.....	15
Panne d'un module d'E/S.....	16
Panne d'un directeur	16
Panne d'une unité d'alimentation du moteur.....	17
Panne d'un ventilateur du moteur	17
Panne d'un sous-réseau IP intracluster	17
Panne d'un switch Fibre Channel intracluster.....	17
Panne d'un moteur VPLEX	17
Panne d'une unité d'alimentation de secours	17
Panne d'un lien intercluster.....	17
Panne du volume de métadonnées	18
Défaillance du journal des zones corrompues	18
Panne du serveur d'administration.....	18
Panne d'un onduleur	19
Panne des clusters VPLEX.....	19
Pannes de l'hôte	19
Pannes des datacenters	19
Conclusion	20
Références	20

Résumé analytique

La famille EMC® VPLEX™ est la solution nouvelle génération en matière de mobilité des données et d'accès aux informations dans et entre les datacenters. C'est la première plate-forme au monde permettant à la fois la fédération locale et la fédération distribuée.

La *fédération locale* permet la coopération transparente des éléments physiques à l'intérieur d'un site. La *fédération distribuée* étend l'accès entre deux sites distants. VPLEX est une solution de fédération entre le *stockage EMC* et le *stockage tiers*.

La solution VPLEX complète l'infrastructure de *stockage virtuel* d'EMC. Elle fournit une couche qui prend en charge le stockage virtuel entre les ordinateurs hôtes qui exécutent les applications du datacenter et les baies de stockage qui fournissent le stockage physique utilisé par lesdites applications.

Dans ce livre blanc, nous étudions la façon dont VPLEX peut être utilisé pour ajouter des niveaux de résilience accrus à des applications exécutées dans et entre les datacenters. Cette capacité est appelée *résilience de la charge de travail*.

Introduction

Ce livre blanc indique comment augmenter la résilience de la charge de travail dans le datacenter grâce à la technologie VPLEX. L'association des techniques et pratiques aujourd'hui courantes et des meilleures pratiques de déploiement de VPLEX permet d'améliorer la résilience de la charge de travail dans le datacenter afin d'assurer la continuité du service en cas d'arrêt des baies et même en cas d'événements ayant un impact important, tels que la maintenance du datacenter ou une coupure d'alimentation. Ces capacités sont introduites par une brève présentation de la technologie VPLEX et des exemples d'utilisation de cette dernière. Vous trouverez ensuite une description du matériel VPLEX, puis une présentation des déploiements courants de cette technologie. Les autres sections du livre décrivent le fonctionnement de VPLEX en cas de divers scénarios de panne pouvant survenir au sein de l'environnement. Des recommandations quant aux meilleures pratiques sont proposées ici pour surmonter tout type de panne et atteindre grâce à VPLEX des niveaux de résilience de charge de travail optimisés dans le datacenter.

Public concerné

Ce livre blanc s'adresse aux architectes et aux administrateurs de stockage qui souhaitent découvrir comment augmenter la résilience de l'infrastructure de stockage du datacenter grâce à VPLEX. Nous supposons qu'ils connaissent les concepts de base relatifs aux baies de stockage, aux réseaux de stockage SAN et à l'infrastructure des serveurs.

Présentation de la technologie VPLEX

EMC VPLEX offre une architecture nouvelle, fruit de plus de 20 ans d'expertise EMC en matière de conception, de mise en oeuvre et de perfectionnement de solutions distribuées de protection des données et de cache intelligent pour les entreprises.

Fondée sur des processeurs évolutifs et hautement disponibles, la famille EMC VPLEX est conçue pour s'adapter en toute transparence à des configurations de toutes tailles. VPLEX s'inscrit entre les serveurs et les ressources de stockage hétérogène et repose sur une architecture unique en cluster, permettant aux serveurs de différents datacenters d'avoir accès en lecture/écriture aux périphériques de stockage de blocs partagé.

Elle présente des caractéristiques uniques :

- un **matériel de clustering évolutif** qui s'adapte à la croissance de vos activités et offre des niveaux de service prévisibles ;
- une **mise en cache avancée des données** qui repose sur une mémoire cache SDRAM de grande capacité afin d'améliorer les performances et de réduire le temps de latence d'E/S et les conflits d'accès aux baies ;
- une **cohérence de cache distribué** offrant partage, équilibrage et basculement sur incident automatiques des E/S à travers tout le cluster ;
- une **vue cohérente** d'une ou plusieurs LUN à travers les clusters VPLEX, à quelques mètres au sein d'un datacenter ou de manière synchrone à distance, donnant lieu à de nouveaux modèles hautement disponibles et à une réaffectation de la charge de travail.

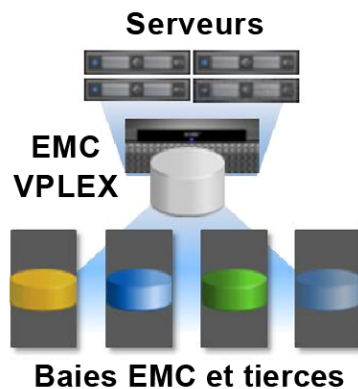


Figure 1. Système EMC VPLEX capable de fédérer des ressources de stockage hétérogène

EMC AccessAnywhere™, disponible avec VPLEX, constitue une avancée considérable puisqu'elle permet aux données d'être partagées, accessibles et réaffectées à distance. **EMC GeoSynchrony™** est le système d'exploitation de VPLEX.

La famille VPLEX comprend deux produits : VPLEX Local et VPLEX Metro.

- **VPLEX Local** garantit une gestion simplifiée et une mobilité des données sans interruption de service entre des baies hétérogènes.
- **VPLEX Metro** fournit un accès aux données et une excellente mobilité entre deux clusters VPLEX sur des distances synchrones.



Figure 2. L'offre de la famille EMC VPLEX et ses limites architecturales

Tirant profit d'une architecture unique totalement évolutive et combinée, la mise en cache avancée des données et la cohérence de cache distribué de la famille VPLEX assurent la résilience, le partage automatique et l'équilibrage de la charge de travail, ainsi qu'un basculement sur incident des domaines de stockage, et permettent un accès aux données locales et distantes avec des niveaux de service prévisibles.

VPLEX Local assure la fédération locale. VPLEX Metro assure la fédération distribuée et étend l'accès entre deux sites à distance synchrone. VPLEX Metro utilise la technologie AccessAnywhere pour permettre aux données d'être partagées, accessibles et réaffectées à distance.

La combinaison d'un datacenter virtualisé et de la solution EMC VPLEX offre aux clients des méthodes entièrement nouvelles pour résoudre leurs problèmes et introduit de nouveaux modèles de technologie informatique. Plus spécialement, les clients peuvent :

- déplacer des applications virtualisées entre les datacenters ;
- permettre l'équilibrage et la réaffectation de la charge de travail entre les sites ;
- consolider des datacenters et assurer une disponibilité 24x7 des services informatiques.

L'architecture en cluster d'EMC VPLEX

VPLEX utilise une architecture nouvelle, en cluster, qui permet aux clients d'éliminer les barrières physiques des datacenters et aux serveurs de différents datacenters d'avoir accès en lecture/écriture aux périphériques de stockage de blocs partagé.

Une configuration VPLEX Local est définie par un, deux ou quatre moteurs VPLEX, intégrés dans un seul cluster grâce à des interconnexions de fabric entre les moteurs totalement redondantes. Grâce à l'interconnexion du cluster, il est possible d'ajouter en ligne des moteurs VPLEX, garantissant ainsi une évolutivité exceptionnelle pour les configurations VPLEX Local et VPLEX Metro. La connectivité entre les moteurs de cluster VPLEX et à travers les configurations VPLEX Metro est entièrement redondante, ce qui assure la protection contre les points uniques de défaillance.

Un cluster VPLEX peut évoluer de façon combinée, grâce à l'ajout de moteurs supplémentaires, et de façon totale, en connectant des clusters sur un VPLEX Metro (deux clusters VPLEX connectés à distance moyenne). VPLEX Metro contribue à la réaffectation et au partage transparents de charges de travail (y compris d'hôtes virtualisés), mais aussi à la consolidation des datacenters et à l'optimisation de l'utilisation des ressources entre datacenters. La mobilité des données sans interruption de service est également assurée, de même que la gestion du stockage hétérogène et la disponibilité optimale des applications. VPLEX Metro prend en charge jusqu'à deux clusters, pouvant se trouver dans le même datacenter, mais sur deux sites différents à distance synchrone (environ 100 kilomètres).

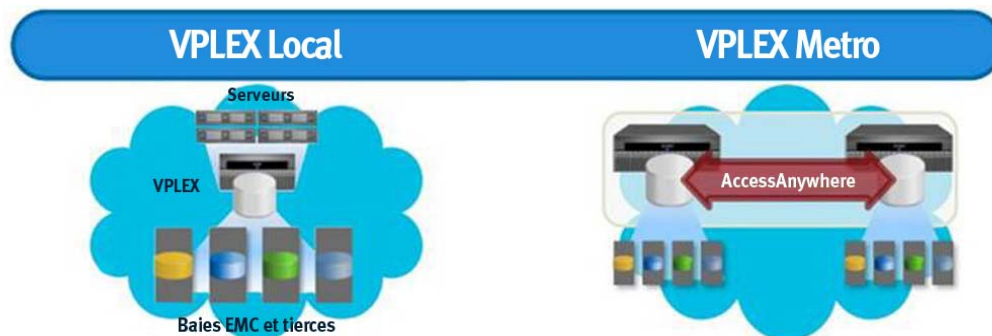


Figure 3. Fédération locale et distribuée avec EMC VPLEX Local et VPLEX Metro

EMC VPLEX répond aux attentes des clients en termes de disponibilité dans le domaine du stockage haut de gamme. Une disponibilité optimale ne se résume pas à une simple redondance. Cela implique des opérations et des mises à niveau sans interruption de service, le système restant « toujours en ligne ». EMC VPLEX fournit :

- AccessAnywhere, avec une connectivité complète des ressources entre les clusters et les configurations Metro-Plex ;
- des options de mobilité et de migration des données entre des baies de stockage hétérogènes ;
- la possibilité de maintenir les niveaux de service et les fonctionnalités pendant la consolidation ;
- un contrôle simplifié pour le provisionnement dans des environnements complexes ;
- un équilibrage de la charge dynamique des données entre les baies de stockage.

Virtualisation des périphériques avec VPLEX

Grâce aux capacités de virtualisation des périphériques dans VPLEX, les périphériques de stockage peuvent être définis en tant que *volumes de stockage* dans le VPLEX, où ils peuvent être divisés ou encapsulés en tant que *domaines* et utilisés pour former des *périphériques* composites présentés aux hôtes en tant que *volumes virtuels*. VPLEX prend en charge plusieurs types de transformation de périphériques virtuels en périphériques physiques, notamment :

- **Extensions** : une extension est un périphérique qui forme une suite de pages contiguës d'un volume de stockage. Une extension peut encapsuler totalement un volume de stockage ou désigner une tranche d'un volume de stockage. Une page VPLEX 4.0 contient 4 Ko de stockage.
- **RAID 0** : un périphérique RAID 0 est l'agrégation d'au moins deux périphériques sur lesquels les pages logiques de stockage sont réparties par bandes afin d'augmenter la performance d'E/S en distribuant les données sur plusieurs piles.
- **RAID-C** : un périphérique RAID-C est l'agrégation d'au moins deux périphériques concaténés logiquement pour former un périphérique de stockage plus important.
- **RAID 1** : un périphérique RAID 1 utilise au moins deux périphériques locaux de même taille pour former un miroir. Chacun de ces périphériques fournit une copie intégrale des données et écrit sur le périphérique RAID 1, chaque tronçon du miroir contenant les données.
- **DR-1** : un périphérique DR-1 est un périphérique RAID 1 distribué. Seule différence par rapport au périphérique RAID 1 : les tronçons du périphérique en miroir sont fournis par différents clusters d'un système VPLEX.

En plus de ses capacités de virtualisation de périphériques, le système VPLEX 4.0 assure la *cohérence de cache distribué*. Le cache VPLEX est maintenu de façon globale et systématique entre les différents directeurs de système. Ensemble, les caches des directeurs entretiennent l'illusion et se comportent comme si chaque volume virtuel était un disque unique alors qu'en réalité, les données du volume peuvent être réparties sur divers périphériques et distribuées entre différents datacenters, et être accessibles depuis plusieurs datacenters.

Présentation du matériel VPLEX

Comme expliqué dans la section précédente, chaque système VPLEX 4.0 est muni d'un ou deux clusters VPLEX, eux-mêmes dotés d'un, deux ou quatre moteurs. Un moteur VPLEX est un châssis composé de deux directeurs, des unités d'alimentation redondantes, de ventilateurs, de modules d'E/S et de modules d'administration. Les directeurs sont les composants phares du système. Ils sont chargés de traiter les demandes d'E/S provenant des hôtes, de mettre les données dans le cache distribué et de les maintenir, de convertir les E/S virtuelles en E/S physiques et d'interagir avec les baies de stockage pour gérer les E/S.

La Figure 4 illustre un VPLEX Metro doté de deux clusters à quatre moteurs.

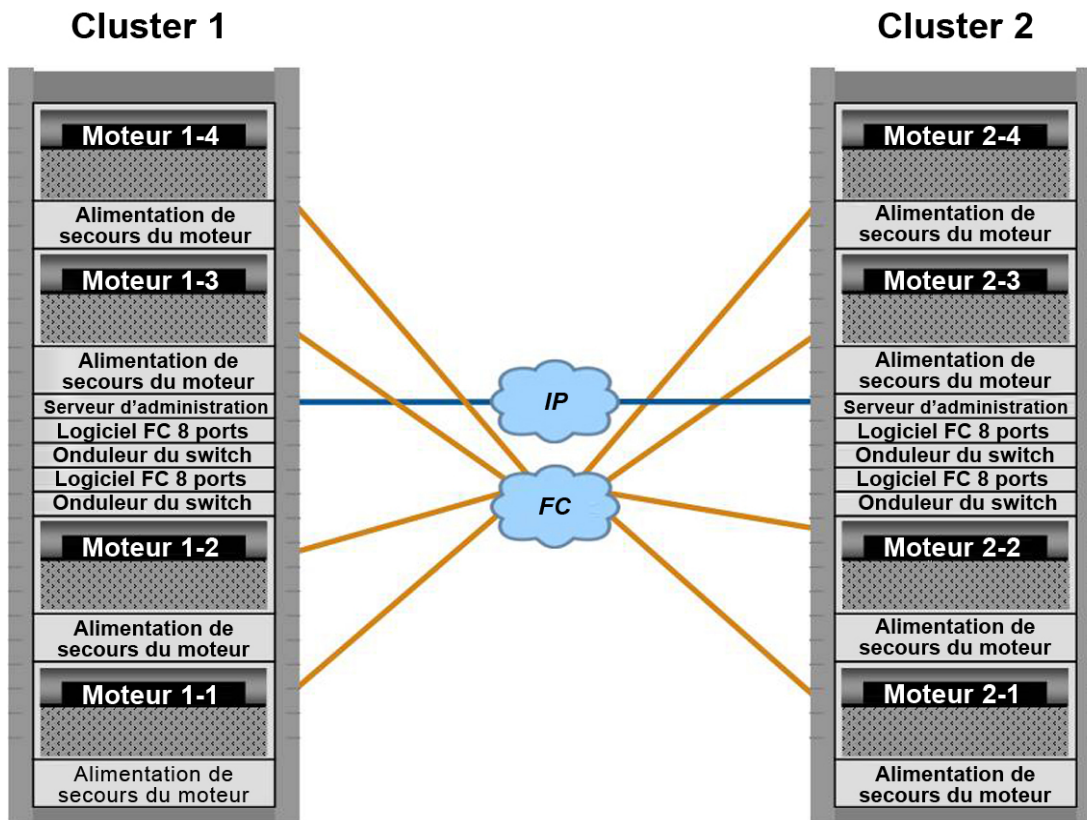


Figure 4. VPLEX Metro avec deux grands clusters

La Figure 5 est un schéma illustrant un moteur VPLEX. Ce schéma illustre les 12 modules d'E/S pris en charge par le moteur, six modules étant affectés à chaque directeur. Chaque directeur est doté de deux modules d'E/S Fibre Channel à quatre ports 8 Gbit/s utilisés pour la connectivité SAN front-end (hôte) et de deux modules d'E/S Fibre Channel à quatre ports 8 Gbit/s utilisés pour la connectivité SAN back-end (baie de stockage). Chacun de ces modules dispose d'une bande passante PCI d'une capacité de 10 Gbit/s connectée au CPU de leur directeur respectif. Un cinquième module d'E/S dispose d'une connectivité Fibre Channel à deux ports 8 Gbit/s pour la communication intracluster et d'une connectivité Fibre Channel à deux ports 8 Gbit/s pour la communication intercluster. Le sixième module d'E/S, actuellement inutilisé, est doté de quatre ports Ethernet à 1 Gbit/s.

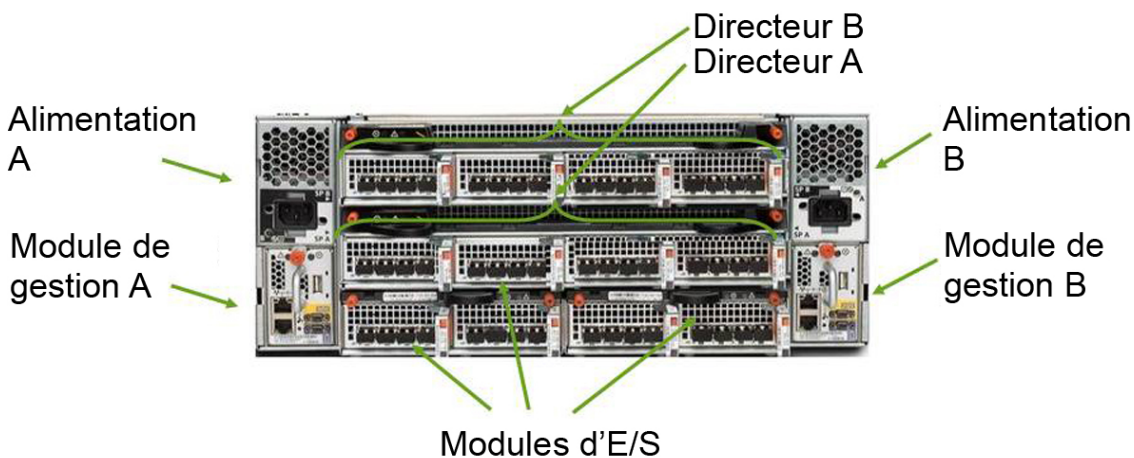


Figure 5. Un moteur VPLEX et ses composants

Le moteur héberge deux unités d'alimentation redondantes, chacune étant capable de fournir l'alimentation nécessaire au châssis. Les modules de gestion redondants fournissent une connectivité IP aux directeurs depuis le serveur de gestion fourni avec chaque cluster. Deux sous-réseaux IP privés fournissent une connectivité IP redondante entre les directeurs d'un cluster et le serveur de gestion du cluster. Quatre ventilateurs redondants assurent la ventilation du châssis du moteur et prennent en charge un modèle de configuration 3+1 qui fournit une ventilation suffisante en cas de panne d'un des ventilateurs.

Chaque moteur est doté d'une alimentation de secours redondante lui permettant de fonctionner en cas de coupure d'alimentation temporaire (cinq minutes maximum).

Les clusters composés de deux moteurs ou plus sont équipés d'une paire de switches Fibre Channel qui fournissent une connectivité Fibre Channel redondante pour la communication intracluster entre les directeurs. Chaque switch Fibre Channel est muni d'un onduleur dédié, qui fournit une alimentation de secours en cas de coupure d'alimentation temporaire.

Présentation du déploiement

VPLEX prend en charge plusieurs modèles de déploiement permettant de répondre à divers besoins. Les prochaines sections décrivent ces différents modèles et les situations dans lesquelles ils doivent être utilisés.

Déploiement VPLEX Local

La Figure 6 illustre un modèle de déploiement type d'un système VPLEX Local. Les systèmes VPLEX Local s'adaptent à des configurations de toutes tailles composées d'un, deux ou quatre moteurs, proposant respectivement des systèmes à deux, quatre ou huit directeurs.

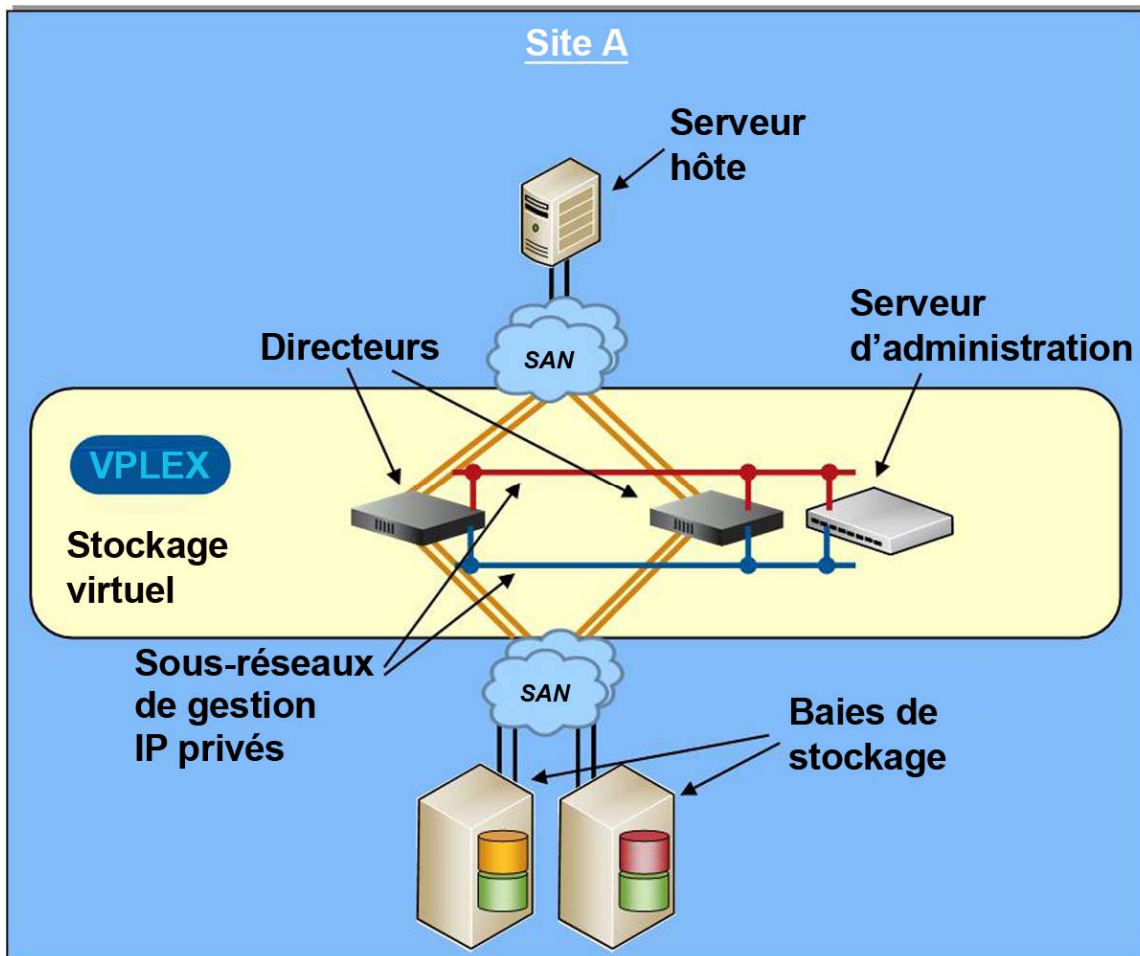


Figure 6. Exemple de déploiement VPLEX Local de petite taille

Cas d'utilisation d'un déploiement VPLEX Local

Vous pouvez déployer VPLEX Local en vue d'utiliser ses capacités de stockage virtuel, telles que la réaffectation de la charge de travail, la résilience de la charge de travail et la gestion du stockage simplifiée, au sein d'un datacenter unique, lorsque la capacité d'évolution de VPLEX Local est suffisante pour répondre aux besoins de ce datacenter. Si une capacité d'évolution supérieure est requise, envisagez le déploiement d'un VPLEX Metro (évoqué ci-après) ou de plusieurs instances de VPLEX Local.

Déploiement de VPLEX Metro au sein d'un datacenter

La Figure 7 illustre un modèle de déploiement type d'un système VPLEX Metro dans un datacenter. Les systèmes VPLEX Metro sont dotés de deux clusters, chacun étant muni d'un, deux ou quatre moteurs. Dans un déploiement VPLEX Metro, les clusters ne doivent pas nécessairement être munis du même nombre de moteurs. Par exemple, un système VPLEX Metro 2 x 4 est équipé d'un cluster doté de deux moteurs et d'un autre doté de quatre moteurs.

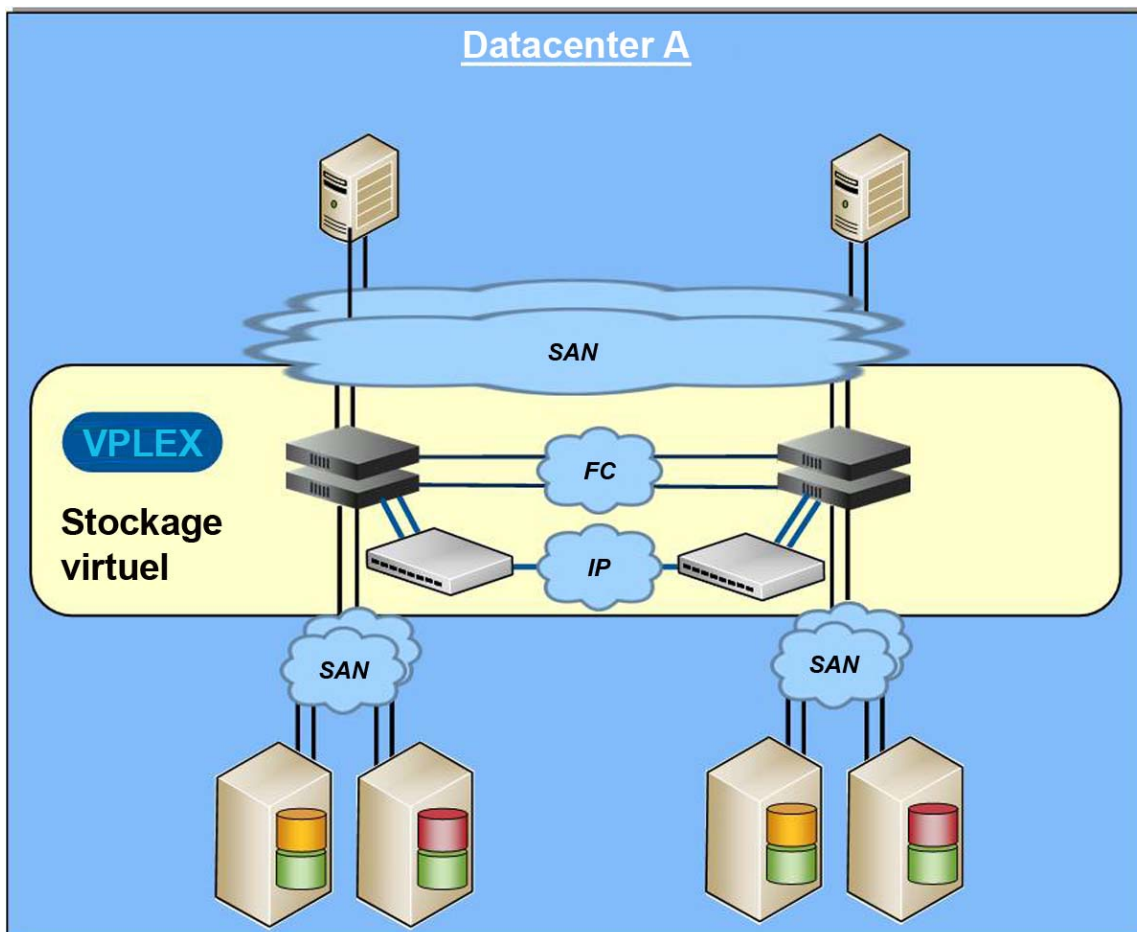


Figure 7. Exemple de déploiement VPlex Metro au sein d'un datacenter

Cas d'utilisation d'un déploiement VPlex Metro au sein d'un datacenter

Vous pouvez déployer VPlex Metro au sein d'un datacenter en vue d'utiliser ses capacités de stockage virtuel, telles que la réaffectation de la charge de travail, la résilience de la charge de travail et la gestion du stockage simplifiée, au sein d'un datacenter unique, lorsqu'une capacité d'évolution supérieure à celle proposée par la solution VPlex Local est requise pour répondre aux besoins de ce datacenter ou lorsqu'une résilience supplémentaire est souhaitée.

Par rapport à VPlex Local, VPlex Metro offre l'avantage supplémentaire suivant en termes de résilience. Les deux clusters d'un VPlex Metro peuvent être séparés par une distance allant jusqu'à 100 km. Le déploiement au sein d'un datacenter est ainsi extrêmement flexible et les deux clusters peuvent être déployés à deux extrémités d'une salle des machines ou à des étages différents pour fournir une meilleure localisation des pannes entre les clusters. En effet, cela permet de placer les clusters dans différentes zones anti-incendie, par exemple, et d'éviter ainsi l'interruption totale du système en cas de panne localisée causée par un incendie circonscrit.

Déploiement VPlex Metro entre des datacenters

La Figure 8 illustre le déploiement d'un système VPlex Metro entre deux datacenters. Ce déploiement est semblable à celui illustré dans la section « Déploiement de VPlex Metro au sein d'un datacenter » excepté que, dans le cas présent, les clusters sont placés dans des datacenters séparés. Cela signifie que des hôtes distincts se connectent à chaque cluster. Les applications en cluster peuvent être dotées, par exemple, d'un ensemble de serveurs d'applications déployés dans le datacenter A et d'un autre ensemble déployé dans le

datacenter B pour une résilience et une réaffectation de la charge de travail accrues. Comme expliqué dans la section précédente, il est important de comprendre qu'une panne totale du site ou du cluster principal d'un volume distribué nécessite la reprise manuelle des E/S sur le site secondaire.

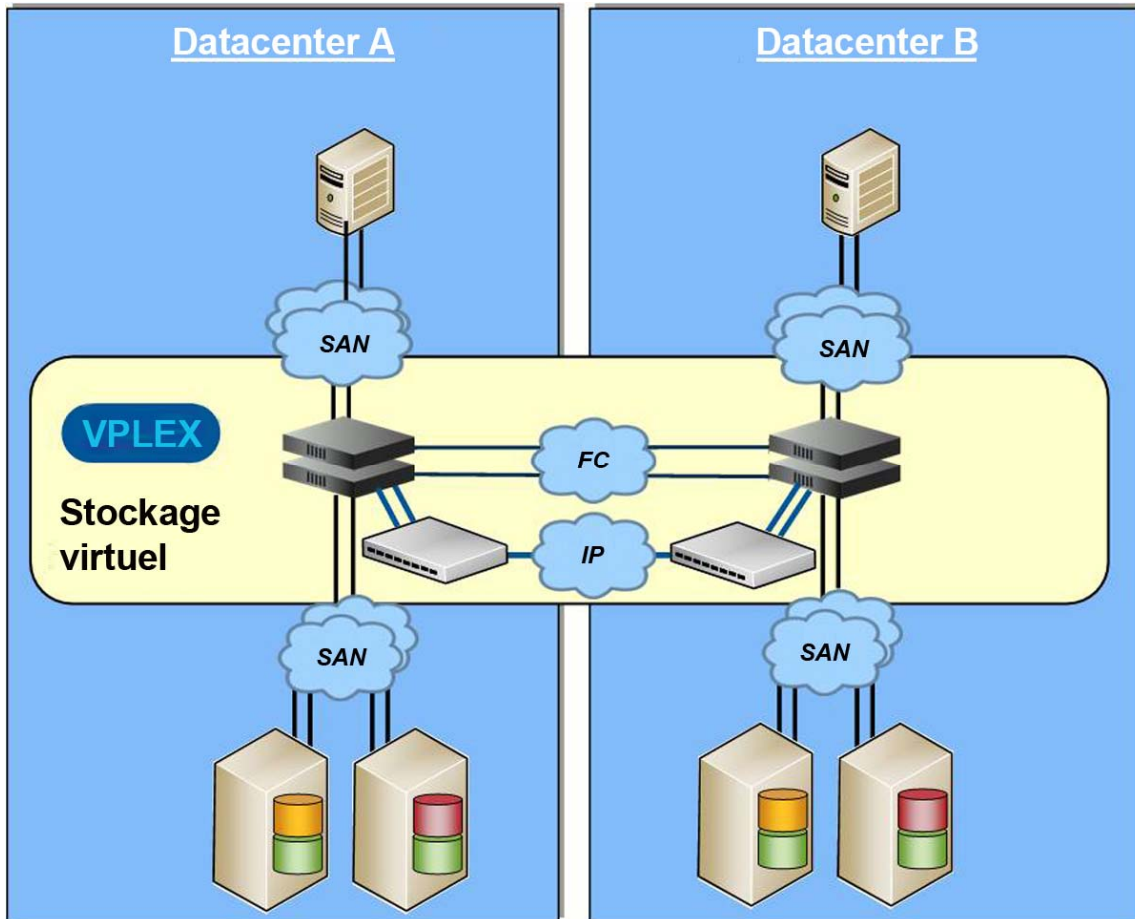


Figure 8. Exemple de déploiement VPlex Metro entre des datacenters

Cas d'utilisation d'un déploiement VPlex Metro entre des datacenters

Vous pouvez déployer VPlex Metro entre deux datacenters lorsque vous souhaitez bénéficier de la résilience supplémentaire de la charge de travail offerte par la présence des données d'une application dans ces deux datacenters. Ce déploiement est également recommandé dans les cas suivants :

- lorsque les applications d'un datacenter veulent accéder aux données de l'autre datacenter ;
- lorsqu'une application souhaite redistribuer les charges de travail entre les deux datacenters ;
- lorsqu'un datacenter ne dispose plus d'espace disponible, n'est plus alimenté ou n'est plus ventilé.

Résilience de la charge de travail

Les prochaines sections sont consacrées à l'étude des différentes pannes pouvant survenir dans un datacenter et à l'ajout possible d'une résilience pour les applications grâce à VPlex, permettant ainsi d'éviter toute incidence sur la charge de travail qui leur est associée. Les types de panne et d'événement étudiés sont les suivants :

- arrêts des baies de stockage (planifiés et non planifiés) ;
- pannes du réseau SAN ;
- défaillance des composants VPlex ;

- panne des clusters VPLEX ;
- pannes de l'hôte ;
- pannes des datacenters.

Arrêts des baies de stockage

Pour parer à des arrêts planifiés et non planifiés des baies de stockage, VPLEX offre la possibilité de mettre en miroir les données d'un volume virtuel entre deux volumes¹ de stockage ou plus à l'aide d'un périphérique RAID 1. La Figure 9 illustre un volume virtuel mis en miroir entre deux baies. En cas d'arrêt planifié ou non planifié d'une baie, le système VPLEX poursuit le traitement des E/S sur le tronçon du miroir encore en fonctionnement. Lors de la restauration du volume de stockage défaillant, les données du volume restant sont resynchronisées vers le tronçon restauré.

Meilleures pratiques

- Il est conseillé de mettre en miroir les données importantes sur au moins deux volumes de stockage gérés par des baies séparées.
- Pour une performance optimale, ces volumes de stockage doivent être configurés de façon identique et gérés par le même type de baie.

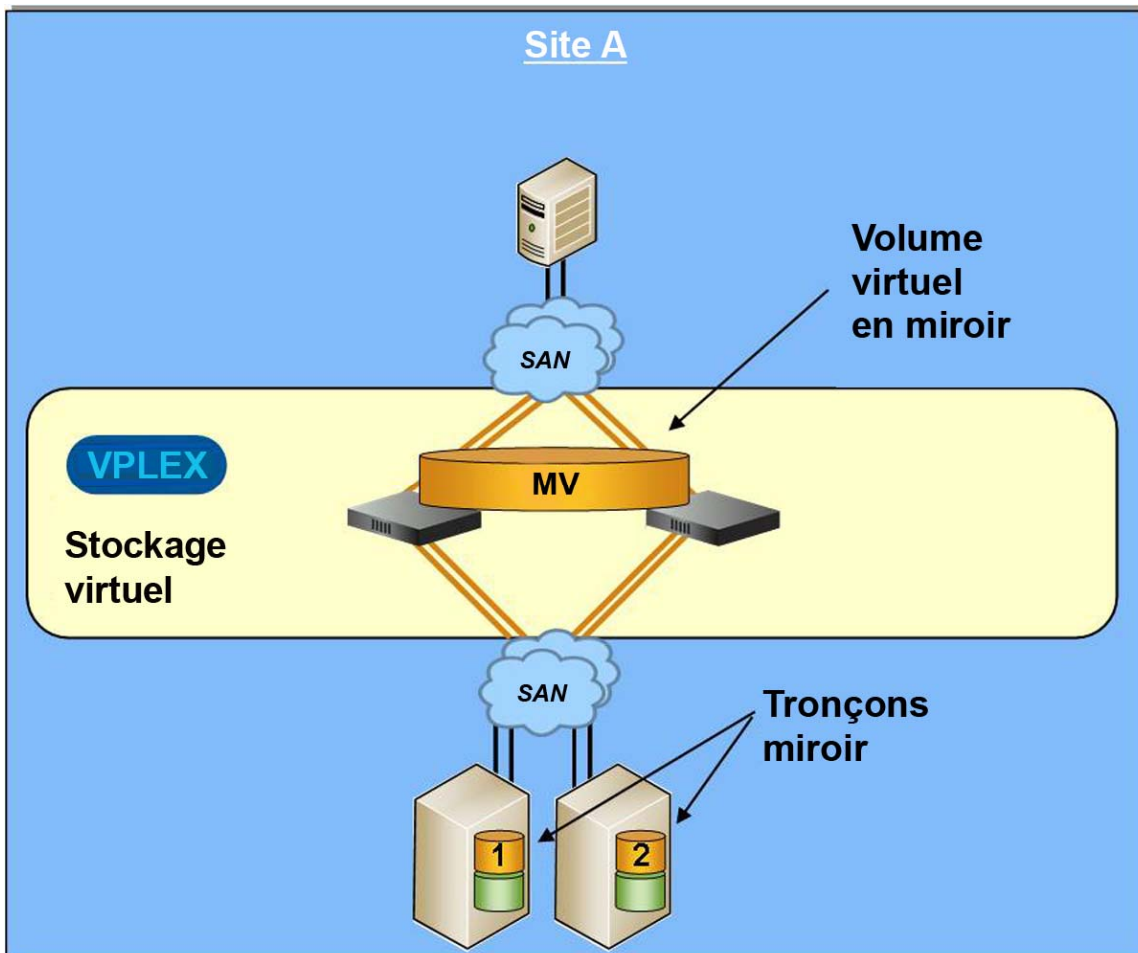


Figure 9. Exemple d'utilisation de la mise en miroir RAID 1 pour se prémunir contre un arrêt des baies

¹ Le système prend en charge jusqu'à huit tronçons par volume miroir.

Pannes du réseau SAN

Lorsqu'une paire de fabrics Fibre Channel redondants est utilisée avec VPLEX, les directeurs VPLEX doivent être connectés aux deux fabrics pour la connectivité front-end (côté hôte), de même que pour la connectivité back-end (côté baie de stockage). Un tel déploiement, combiné à l'isolation des fabrics, permet au système VPLEX de fonctionner même en cas de panne d'un fabric entier et permet au système d'assurer un accès continu aux données dans le cas d'une telle panne. Les hôtes doivent également être connectés aux deux fabrics et utiliser des logiciels de multipathing pour assurer un accès continu aux données dans le cas d'une telle panne. La Figure 10 illustre une bonne pratique pour le déploiement de deux fabrics.

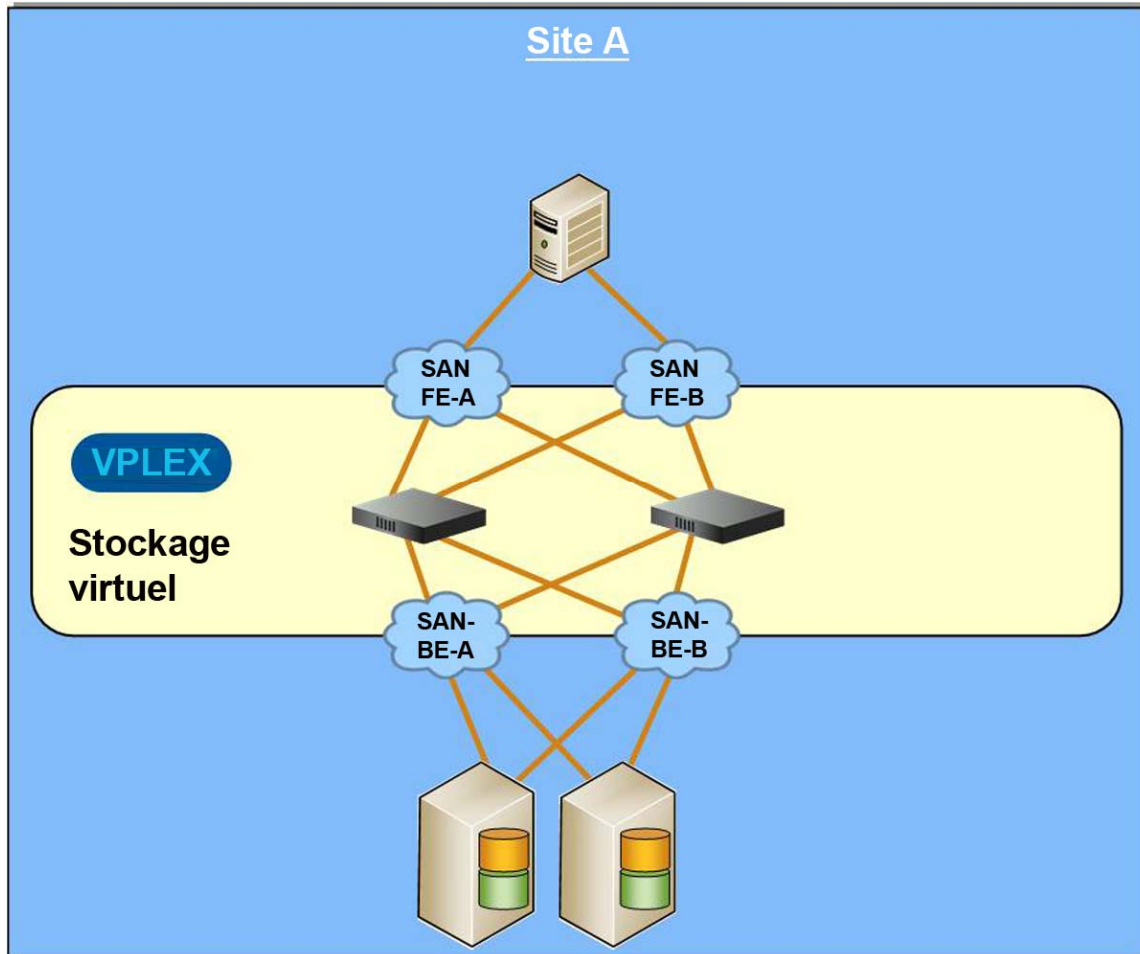


Figure 10. Déploiement recommandé de deux fabrics

Bonne pratique

- Il est recommandé de connecter les modules d'E/S à des fabrics redondants. Par exemple, dans le cas d'un déploiement avec les fabrics A et B, il est recommandé de connecter les ports d'un directeur tel qu'illustré à la Dans des clusters VPLEX dotés de deux moteurs ou plus, l'éventualité peu probable d'une panne de moteur entraîne l'interruption de service Figure 11.

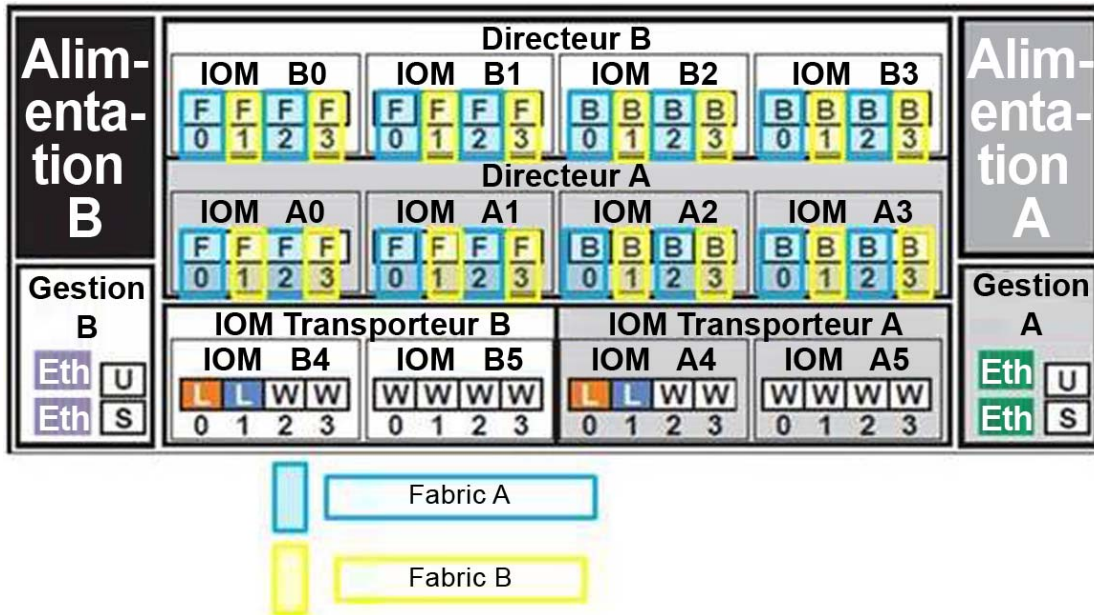


Figure 11. Affectations de fabric recommandées pour les ports FE et BE

Défaillance des composants VPLEX

Tous les composants importants d'un système VPLEX utilisent au minimum une redondance par paires pour optimiser l'accès aux données. Cette section décrit le mode de gestion des défaillances de composants VPLEX et les meilleures pratiques à suivre pour permettre aux applications de tolérer ces pannes.

Toutes les défaillances de composants survenant dans un système VPLEX sont consignées par le biais d'événements transmis automatiquement à EMC Service Center pour garantir une réponse et un dépannage rapides.

Défaillance de port Fibre Channel

Toutes les communications VPLEX passent par des chemins redondants, évitant ainsi toute interruption en cas de défaillance des ports. Cette redondance permet aux logiciels de multipathing des serveurs hôtes de retransmettre et de rediriger les E/S en contournant le chemin touché par une défaillance des ports ou d'autres événements dans le réseau SAN.

VPLEX utilise sa propre logique de multipathing pour maintenir la redondance des chemins d'accès au stockage back-end depuis chaque directeur. Cela permet à VPLEX de fonctionner même en cas de défaillance des ports back-end du VPLEX ainsi que des fabrics back-end et des ports des baies assurant la connexion du stockage physique à VPLEX.

Les émetteurs-récepteurs enfichables à faible encombrement (SFP) utilisés pour la connectivité VPLEX sont des modules prêts à l'emploi remplaçables sur site.

Meilleures pratiques

- Assurez-vous qu'un chemin relie chaque hôte à au moins un port front-end sur le directeur A et à au moins un port front-end sur le directeur B. Si le système VPLEX est doté de deux moteurs ou plus, assurez-vous que l'hôte dispose au moins d'un chemin « côté A » dans un moteur et d'au moins un chemin « côté B » dans un moteur distinct. Pour une disponibilité optimale, chaque hôte peut disposer d'un chemin d'accès à au moins un port front-end sur chaque directeur.
- Utilisez le logiciel de multipathing sur les serveurs hôtes pour garantir une réponse rapide et l'absence d'interruption des E/S en cas de défaillance d'un chemin.

-
- Assurez-vous que chaque hôte dispose d'un chemin d'accès à chaque volume virtuel par le biais de chaque fabric.
 - Assurez-vous que le mappage et le masquage de LUN de chaque volume de stockage présenté à VPLEX depuis une baie de stockage présentent les volumes sortants d'au moins deux ports de la baie sur au moins deux fabrics différents et se connectent à au moins deux ports différents desservis par deux modules d'E/S back-end distincts pour chaque directeur au sein d'un cluster VPLEX.
 - Assurez-vous que la segmentation du fabric fournit aux hôtes un accès redondant aux ports front-end VPLEX et assure à VPLEX l'accès redondant aux ports des baies.

Panne d'un module d'E/S

Les modules d'E/S au sein de VPLEX remplissent des fonctions spécifiques. Chaque directeur VPLEX est doté de deux modules d'E/S front-end, de deux modules d'E/S back-end et d'un module d'E/S COM utilisé pour la connectivité intracluster et intercluster. Tous les modules d'E/S sont des unités prêtes à l'emploi remplaçables sur site. Les sections suivantes décrivent le fonctionnement du système et les meilleures pratiques pour optimiser la disponibilité en cas de panne.

Module d'E/S FE

En cas de panne d'un module d'E/S FE, tous les chemins connectés à ce module d'E/S sont interrompus et cessent de fonctionner. Vous devez respecter les meilleures pratiques énumérées à la page 15 pour vous assurer que les hôtes disposent d'un chemin redondant vers leurs données.

Lors du retrait et du remplacement d'un module d'E/S, le directeur concerné est réinitialisé.

Module d'E/S BE

En cas de panne d'un module d'E/S BE, tous les chemins connectés à ce module d'E/S sont interrompus et cessent de fonctionner. Vous devez respecter les meilleures pratiques énumérées à la page 15 pour vous assurer que chaque directeur dispose d'un chemin d'accès redondant à chaque volume de stockage par le biais d'un module d'E/S distinct.

Lors du retrait et du remplacement d'un module d'E/S, le directeur concerné est réinitialisé.

Module d'E/S COM

En cas de panne du module d'E/S COM d'un directeur, ce dernier est réinitialisé et tous les services fournis depuis ce directeur sont interrompus. Vous devez respecter les meilleures pratiques énumérées à la page 15 pour vous assurer que chaque hôte dispose d'un accès redondant à son volume de stockage virtuel par le biais de plusieurs directeurs, de sorte que l'hôte a toujours accès à son stockage après la réinitialisation d'un directeur.

Lors du retrait et du remplacement d'un module d'E/S, le directeur concerné est réinitialisé.

Panne d'un directeur

La panne d'un directeur entraîne la perte de tous les services fournis par ce dernier. Chaque moteur VPLEX est muni d'une paire de directeurs à des fins de redondance. Les clusters VPLEX dotés de deux moteurs ou plus bénéficient d'une redondance supplémentaire fournie par les directeurs supplémentaires. Chaque directeur au sein d'un cluster peut présenter le même stockage. Les meilleures pratiques décrites à la page 15 permettent à un hôte de continuer à fonctionner en cas de panne d'un directeur grâce à la création de chemins d'accès redondants à leur stockage virtuel par le biais de ports fournis par différents directeurs. La combinaison de logiciels de multipathing sur les hôtes et de chemins redondants par le biais de différents directeurs du système VPLEX permet à l'hôte de fonctionner malgré la perte d'un directeur.

Dans un système à moteurs multiples, un hôte peut conserver l'accès à ses données dans l'éventualité peu probable où plusieurs directeurs tomberaient en panne grâce aux chemins d'accès à son stockage virtuel fournis par chaque directeur du système.

Tous les directeurs sont des unités prêtes à l'emploi remplaçables sur site.

Panne d'une unité d'alimentation du moteur

Les unités d'alimentation du moteur VPLEX sont entièrement redondantes. En cas de panne d'une unité d'alimentation, aucun service ni aucune fonction ne sont interrompus.

Toutes les alimentations sont des unités prêtes à l'emploi remplaçables sur site qui peuvent être retirées et remplacées sans interruption du système.

Panne d'un ventilateur du moteur

Les modules de ventilation des moteurs VPLEX sont entièrement redondants. En cas de panne d'un ventilateur unique, le service n'est pas interrompu. Chaque moteur est doté de quatre ventilateurs. En cas de panne d'un ventilateur, les trois ventilateurs restants garantissent la ventilation suffisante du système. En cas de panne de deux ventilateurs, le moteur s'arrête automatiquement pour éviter tout dommage lié à la surchauffe.

Tous les modules de ventilation sont des unités prêtes à l'emploi remplaçables sur site qui peuvent être retirées et remplacées sans interruption du système.

Panne d'un sous-réseau IP intracluster

Chaque cluster VPLEX dispose d'une paire de sous-réseaux IP locaux privés qui relient les directeurs au serveur de gestion. Ces sous-réseaux sont utilisés pour le trafic de gestion et pour éviter le partitionnement intracluster. En cas de perte d'un lien sur l'un de ces sous-réseaux, certains membres ne pourront plus communiquer avec d'autres membres sur ce sous-réseau. Grâce à la présence du sous-réseau redondant, le service et la gestion ne sont pas interrompus.

Panne d'un switch Fibre Channel intracluster

Chaque cluster VPLEX doté de deux moteurs ou plus utilise une paire de switches Fibre Channel dédiés pour la communication intracluster entre les directeurs au sein du cluster. Deux fabrics Fibre Channel redondants sont créés, chaque switch étant affecté à un fabric différent. En cas de panne d'un switch Fibre Channel, le traitement et le service ne sont pas interrompus.

Panne d'un moteur VPLEX

Dans des clusters VPLEX dotés de deux moteurs ou plus, l'éventualité peu probable d'une panne de moteur entraîne l'interruption de service des directeurs au sein dudit moteur, mais les volumes virtuels desservis par les directeurs situés dans les autres moteurs demeurent disponibles. La section des « Meilleures pratiques » à la page 15 décrit les meilleures pratiques pour placer des chemins d'accès redondants vers un volume virtuel sur les directeurs depuis divers moteurs dans des clusters VPLEX à moteurs multiples.

Panne d'une unité d'alimentation de secours

Chaque moteur VPLEX est pris en charge par une paire d'unités d'alimentation de secours qui permettent au système de fonctionner en cas de coupure d'alimentation temporaire (cinq minutes maximum). Une seule unité d'alimentation de secours suffit à alimenter le moteur connecté. VPLEX dispose d'une paire d'unités d'alimentation de secours à des fins de haute disponibilité.

Chaque unité d'alimentation de secours peut être remplacée sur site, sans interruption des services assurés par le système. Le temps de recharge d'une alimentation de secours est de 5,5 heures maximum et les batteries des alimentations de secours peuvent supporter deux coupures séquentielles de cinq minutes.

Panne d'un lien intercluster

Chaque directeur d'un système VPLEX est doté de deux liens dédiés à la communication intercluster. Chacun de ces liens doit être configuré (par exemple, segmenté) pour fournir des chemins d'accès à

chaque directeur dans le cluster distant. De cette façon, la connectivité intégrale entre les directeurs demeure disponible, même en cas de panne d'un lien. En cas de perte des deux liens du directeur, tous les modules d'E/S intercluster sont suspendus entre les deux clusters afin de respecter l'ordre d'écriture et de s'assurer que le site distant conserve une image récupérable. Le cas échéant, des miroirs distants sont séparés et les règles définies par l'utilisateur, ou *règles de déconnexion*, sont exécutées pour déterminer quel cluster VPLEX doit continuer à autoriser les E/S pour un miroir distant donné. Ces règles peuvent être configurées pour chaque périphérique, permettant ainsi à certains volumes de rester disponibles sur un cluster, tandis que d'autres volumes restent disponibles sur l'autre cluster.

Une fois les liens réparés, les E/S peuvent être restaurées et les tâches de resynchronisation lancées pour restaurer les miroirs distants. Ces actions s'effectuent automatiquement. Les volumes peuvent également être configurés de façon à nécessiter la reprise manuelle des E/S si la coordination avec les actions du serveur est requise. Les opérations d'E/S des périphériques peuvent être effectuées immédiatement sans attendre la fin des tâches de resynchronisation.

Panne du volume de métadonnées

VPLEX conserve son état de configuration, ou *métadonnées*, sur les volumes de stockage fournis par les baies de stockage du réseau SAN. Chaque cluster VPLEX gère ses propres métadonnées ; celles-ci décrivent les informations relatives à la configuration locale dudit cluster, ainsi que toutes les informations relatives à la configuration distribuée partagées entre les clusters. Il est fortement recommandé de configurer le volume des métadonnées de chaque cluster avec des volumes de stockage back-end multiples fournis par différentes baies de stockage du même type. Les capacités de protection des données qu'offrent ces baies de stockage, telles que RAID 1 et RAID 5, doivent être utilisées pour garantir l'intégrité des métadonnées du système. En outre, il est fortement recommandé d'effectuer des copies de sauvegarde des métadonnées en cas de modification de la configuration du système.

VPLEX utilise ces métadonnées persistantes lors du démarrage complet du système et charge les informations de configuration sur chaque directeur. Lorsque la configuration du système est modifiée, les changements apportés sont inscrits sur le volume des métadonnées. En cas d'interruption de l'accès au volume des métadonnées, les directeurs VPLEX continuent à fournir leurs services de virtualisation grâce à la copie des informations de configuration en mémoire. Si le stockage qui prend en charge le périphérique des métadonnées demeure indisponible, un nouveau périphérique de métadonnées doit être configuré. Une fois le nouveau périphérique affecté et configuré, la copie du périphérique de métadonnées en mémoire conservée par le cluster est enregistrée sur le nouveau périphérique de métadonnées.

Vous ne pouvez pas modifier la configuration lorsque l'accès au périphérique de métadonnées persistantes est indisponible.

Défaillance du journal des zones corrompues

VPLEX Metro utilise un journal des zones corrompues pour enregistrer les informations indiquant les zones d'un miroir distribué séparé ayant été mises à jour alors qu'un tronçon de miroir était déconnecté. Les informations relatives à chaque tronçon de miroir déconnecté sont conservées. Si ce volume devient inaccessible, les directeurs enregistrent le tronçon entier comme étant obsolète et demandent une resynchronisation totale de ce tronçon de volume une fois ce dernier rattaché au miroir.

Panne du serveur d'administration

Chaque cluster VPLEX dispose d'un serveur d'administration dédié qui fournit aux directeurs un accès aux opérations de gestion et prend en charge la connectivité de gestion pour l'accès à distance au cluster homologue dans un environnement VPLEX Metro. Puisque le traitement des E/S des directeurs VPLEX ne dépend pas des serveurs d'administration, la panne d'un serveur d'administration n'entraîne pas l'interruption des services de traitement et de virtualisation des E/S fournis par VPLEX.

Panne d'un onduleur

Dans des clusters VPLEX dotés de deux moteurs ou plus, une paire de switches Fibre Channel prend en charge la communication intracluster entre les directeurs de ces moteurs. Chaque switch est muni d'un onduleur dédié qui assure l'alimentation en cas de coupure d'alimentation temporaire. Les onduleurs permettent aux switches Fibre Channel de continuer à fonctionner pendant cinq minutes (maximum) après une coupure d'alimentation. L'onduleur en bas de l'armoire fournit également une alimentation de secours pour le serveur d'administration.

Panne des clusters VPLEX

VPLEX Metro prend en charge deux types de périphérique distribué : *les volumes virtuels distribués à distance moyenne* et *les volumes virtuels distants*. Les volumes virtuels distribués à distance moyenne fournissent des copies synchronisées (miroirs) des données du volume dans chaque cluster. Le volume mis en miroir apparaît et se comporte comme un volume unique. Il agit de la même manière qu'un volume virtuel utilisant un périphérique RAID 1, avec toutefois une valeur ajoutée puisque chaque cluster conserve une copie des données. Les volumes virtuels distants fournissent l'accès à un volume virtuel dont les données résident dans un cluster. Les volumes virtuels distants, tout comme les volumes distribués à distance moyenne, sont capables d'exploiter le cache cohérent distribué VPLEX et ses algorithmes de lecture préalable pour des performances supérieures à une solution d'extension SAN.

Pour chaque volume virtuel distribué à distance moyenne, une règle de déconnexion identifie le cluster au sein d'un système VPLEX Metro qui doit être déconnecté du tronçon de son miroir (mis hors service) en cas de perte de la communication entre les deux clusters. Ces règles définissent un site préféré (ou gagnant) dans le cas où la communication entre les clusters est interrompue. Cette interruption de communication entre les clusters peut avoir deux causes différentes : une panne de lien intra-clusters, tel qu'évoqué dans la section « Panne d'un onduleur d'un lien intercluster » et une panne de cluster. La présente section décrit la panne de cluster. En cas de panne de cluster, les E/S de tout volume virtuel distribué à distance moyenne qui dispose d'une règle de déconnexion ayant identifié le site restant comme le site gagnant restent en service sur le tronçon restant du périphérique. Les E/S sont interrompues pour les volumes dont les règles de déconnexion indiquent que ce site doit être déconnecté en cas d'interruption de la communication. Comme il est impossible de distinguer une panne de cluster d'une panne de lien, ce fonctionnement vise à préserver l'intégrité des données sur les périphériques distribués.

Les volumes virtuels distants peuvent faire l'objet de deux types de panne. Dans le premier cas, si le cluster qui fournit le support physique du volume virtuel tombe en panne, le volume virtuel distant est entièrement inaccessible. Dans le second cas, si le cluster distant tombe en panne (le cluster sans support physique pour ce volume), l'accès au volume virtuel demeure possible depuis le cluster hôte (le cluster contenant les données physiques), mais non depuis le cluster distant.

Pannes de l'hôte

Bien que cette fonctionnalité ne soit pas fournie par VPLEX, le clustering basé sur l'hôte est une technique essentielle pour optimiser la résilience de la charge de travail. Elle permet en effet à une application de continuer à fournir des services en cas de panne de l'hôte grâce à un modèle de traitement actif/actif ou actif/passif. L'infrastructure obtenue, combinée aux capacités de VPLEX susmentionnées, permet d'atteindre des niveaux de disponibilité très élevés.

Pannes des datacenters

Un VPLEX Metro distribué entre deux datacenters peut être utilisé pour éviter la perte de données en cas de panne d'un datacenter grâce à la mise en miroir des données entre deux datacenters. En cas de panne d'un datacenter, ce déploiement peut améliorer l'accès aux données. Pour savoir comment, consultez la section « Panne des clusters VPLEX » (dans le cas où la panne du datacenter est la conséquence de la perte de l'un des clusters VPLEX dans un VPLEX Metro). L'accès aux données demeure disponible pour les volumes virtuels distribués à distance moyenne dont le cluster gagnant est le datacenter restant. Pour les volumes dont le cluster gagnant est le datacenter en panne, l'accès aux données peut être restauré dans

l'autre cluster grâce à l'utilisation d'une commande manuelle permettant de reprendre les E/S interrompues. Combinée à une logique de basculement sur incident pour les clusters de l'hôte, cette technique fournit une infrastructure capable de rétablir rapidement les opérations du service, même en cas d'arrêt non planifié du datacenter.

Certaines pannes de datacenter sont causées par une coupure d'alimentation dans le datacenter. VPLEX utilise des alimentations de secours et des onduleurs pour parer aux coupures d'alimentation temporaires (cinq minutes maximum). Cette technologie doit être combinée à une infrastructure auxiliaire identique pour les hôtes, l'équipement réseau et les baies de stockage afin d'obtenir une solution complète de tolérance des coupures d'alimentation temporaires. En cas de coupure d'alimentation de plus de cinq minutes, VPLEX interrompt les services de virtualisation. Les propriétés de mise en mémoire cache à écriture immédiate de VPLEX garantissent l'écriture des données de l'application sur les baies de stockage back-end avant leur transmission à l'hôte.

Conclusion

VPLEX offre une redondance matérielle et logicielle interne étendue qui assure non seulement la haute disponibilité des services VPLEX, mais offre en outre une résilience accrue de la charge de travail de l'infrastructure environnante. Combinée aux meilleures pratiques en matière de clustering basé sur l'hôte, de multipathing, de redondance de fabric, de protection des supports de stockage et d'alimentation de secours, cette solution offre des bases solides pour permettre au stockage virtuel de répondre efficacement aux exigences de disponibilité du stockage.

Références

Pour plus d'informations sur l'infrastructure de stockage virtuel et les fonctionnalités de VPLEX 4.0, consultez les livres blancs suivants.

- *Meilleures pratiques de mise en oeuvre et de planification pour EMC VPLEX – Notes techniques*
- *Utilisation des plates-formes de virtualisation VMware avec EMC VPLEX – Meilleures pratiques de planification*
- *Déplacement des données sans interruption de service : événements planifiés avec EMC VPLEX - Meilleures pratiques de planification*
- *Activation de VMotion à distance pour Microsoft, Oracle et SAP grâce à VCE Vblock1, EMC Symmetrix VMAX, EMC CLARiiON et EMC VPLEX Metro – Présentation de l'architecture*
- *Implémentation d'EMC VPLEX et Microsoft Hyper-V et SQL Server avec support amélioré du clustering avec basculement – Technologie appliquée*