

EMC XTREMIO : SOLUTION DE CONSOLIDATION HAUTES PERFORMANCES POUR ORACLE

EMC XtremIO, VMware vSphere, base de données Oracle, Oracle RAC

- Optimiser les temps de service du stockage pour les charges applicatives OLTP et OLAP/d'entrepôt décisionnel à débits élevés
- Virtualiser et consolider les instances de base de données avec de meilleures performances
- Réduire le nombre de copies et simplifier l'environnement de test et de développement pour une meilleure efficacité informatique

Solutions EMC

Résumé

Ce livre blanc décrit les avantages en matière de performances et d'utilisation d'une base de données Oracle virtualisée, avec VMware vSphere déployé sur une baie EMC® XtremIO™ 100 % Flash. Il explique également comment la solution optimise la consolidation et la virtualisation des environnements de base de données Oracle.

Juillet 2014

Copyright © 2014 EMC Corporation. Tous droits réservés.

Publié en juillet 2014.

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 ce document 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.

EMC², EMC et le logo EMC sont des marques déposées ou des marques commerciales d'EMC Corporation aux États-Unis et dans d'autres pays. Toutes les autres marques citées dans le présent document sont la propriété de leurs détenteurs respectifs.

Pour obtenir la liste actualisée des noms de produits, consultez la rubrique des marques EMC via le lien Législation, sur <http://france.emc.com>.

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

Référence H12117.1

Table des matières

Résumé analytique	6
Business case	6
Présentation de solution	6
Résultats principaux	7
Introduction	8
Objectif	8
Périmètre	8
Public	8
Présentation de la technologie	9
Présentation	9
EMC XtremIO	9
Voici quelques-unes des fonctions principales d’XtremIO :	11
Performances évolutives	11
Haute disponibilité	11
Réduction des données à la volée	11
Snapshot XtremIO	11
Prise en charge de VMware VAAI	12
VMware vSphere	13
VMware vCenter	13
EMC PowerPath/VE	13
Oracle Database 11g Enterprise Edition	13
Oracle RAC 11gR2	13
Oracle ASM	14
Oracle Clusterware	14
Architecture de la solution	15
Présentation	15
Schéma de l’architecture	15
Ressources matérielles	16
Ressources logicielles	16
Couche de stockage : EMC XtremIO	17
Conception de cluster	17
Réduction des données à la volée	17
Allocation dynamique	18
Protection contre les pannes	18
Évolutivité	18
Opérations sur les métadonnées in-memory	19
XtremIO Management Server	19
Configuration du stockage	19
Conception du stockage	20
Provisionnement des LUN	21

Couche réseau	24
Présentation	24
Bonnes pratiques SAN	24
Bonnes pratiques relatives au réseau IP.....	24
Bonnes pratiques relatives au réseau VMware vSphere	24
Couche de traitement	26
Présentation	26
Ressources de traitement et de stockage	26
Virtualisation	27
Virtualisation du réseau	27
Configuration des modèles de machine virtuelle	27
Autorisation d'accès aux périphériques de stockage	29
Base de données Oracle	30
Présentation	30
Conception du stockage des bases de données OLTP	30
Organisation des groupes de disques ASM pour la base de données OLTP	31
Profil de base de données et de charge applicative OLTP	31
Configuration du groupe de disques ASM pour la base de données d'entrepôt décisionnel	31
Profil de base de données et de charge applicative DW.....	32
Test et validation des performances	33
Présentation	33
Remarques sur les résultats	33
Objectif du test	33
Scénarios de test	34
Tests sur les charges applicatives pour la lecture/écriture de blocs uniques aléatoires SLOB	34
OLTP avec test 100 % requêtes : test portant uniquement sur les requêtes de blocs uniques aléatoires SLOB	34
OLTP avec test 100 % UPDATE : test sur les lectures/écritures de blocs uniques aléatoires SLOB	36
OLTP avec test 75 %/25 % QUERY/UPDATE test sur les blocs uniques aléatoires SLOB	39
Test sur la charge applicative des requêtes d'entrepôt décisionnel.....	41
Méthodologie de test	41
Procédure de test	41
Résultats des tests	41
Test de chargement des données d'entrepôt décisionnel.....	42
Méthodologie de test	42
Procédure de test	42
Résultats des tests	43
Conclusion	44
Résumé.....	44
Conclusions	44

Références	45
Livres blancs	45
Documentation VMware	45
Documentation Oracle	45
Autre documentation	45

Résumé analytique

Business case

Les besoins métiers entraînent la croissance du volume des données collectées et de la vitesse à laquelle elles le sont. Parallèlement, la demande de transformation rapide de ces données en informations sur les opportunités et les risques qu'elles représentent se fait plus pressante.

Des bases de données telles qu'Oracle Database 11g sont utilisées pour prendre en charge les applications critiques. Pour offrir des temps de réponse rapides sur la gamme d'applications, ces bases de données requièrent l'utilisation d'un stockage conçu à la fois pour les E/S transactionnelles avec latence faible et pour les charges applicatives d'analytique à débit élevé.

La virtualisation permet une meilleure consolidation des types de charge applicative de base de données. Souvent, de par la consolidation, les charges applicatives OLTP (Online Transaction Processing) et OLAP (Online Analytical Processing) partagent les mêmes serveurs et le même stockage. Par conséquent, pour des performances optimales, l'infrastructure de stockage sous-jacente doit également être conçue pour gérer les différentes charges applicatives d'une infrastructure consolidée.

La baie EMC® XtremIO™ 100 % Flash gère efficacement les effets de la virtualisation sur les charges applicatives de base de données exigeantes en termes d'E/S, avec des performances d'E/S aléatoires impressionnantes et des temps de latence très faibles et cohérents. Cela s'applique aussi bien aux charges applicatives OLTP qu'OLAP, ainsi qu'à la consolidation de plusieurs charges applicatives sur une plate-forme de stockage commune. XtremIO confère également aux environnements virtualisés une vitesse et une agilité de provisionnement inédites : snapshots peu encombrants, déduplication des copies à la volée, allocation dynamique et provisionnement accéléré via VMware VAAI (VMware vStorage APIs for Array Integration). Il en résulte des économies sur les licences de stockage et de base de données, une simplicité exceptionnelle pour ce qui est de la gestion et du provisionnement du stockage, ainsi que de nouvelles fonctions pour l'analytique en temps réel et les cycles de développement/test.

Présentation de solution

La virtualisation des serveurs de bases de données constitue une stratégie éprouvée pour la consolidation des bases de données, mais elle peut représenter des défis spécifiques. Lors de la consolidation de plusieurs applications sur un nombre d'hôtes physiques réduit, la charge applicative d'E/S peut devenir extrêmement aléatoire pour le stockage back-end, car les machines virtuelles partagent des ressources physiques telles que les adaptateurs HBA. La baie de stockage XtremIO 100 % Flash est conçue pour s'exécuter dans ces environnements virtualisés exigeants et offrir des performances d'E/S aléatoires élevées et cohérentes tout en optimisant les coûts par base de données.

Cette solution démontre les avantages du déploiement d'Oracle Database 11g sur VMware vSphere avec un système de stockage XtremIO.

Les principaux avantages de cette solution sont les suivants :

- consolidation des bases de données Oracle virtualisées et des autres applications sur les mêmes systèmes de stockage et hôtes physiques ;
- disponibilité et performances prévisibles pour les contrats de niveau de service des applications ;

- facilité d'utilisation et planification de la croissance simplifiée avec XtremIO ;
- encombrement du stockage optimal pour l'ensemble de l'environnement de base de données, y compris toutes les copies de production et de test/développement.

Résultats principaux

Dans le cadre de cette solution, XtremIO offre :

- une configuration rapide et simple, avec peu voire aucun réglage du stockage requis ;
- des E/S aléatoires de blocs de taille volumineuse et une bande passante élevée pour un grand entrepôt décisionnel avec plusieurs utilisateurs simultanés ;
- des IOPS (E/S par seconde) accrues et une latence cohérente pour l'évolutivité des charges applicatives OLTP ;
- des performances prévisibles pour les environnements de base de données consolidés et virtualisés avec E/S aléatoires ;
- des économies importantes en termes d'encombrement du stockage via l'utilisation de la déduplication à la volée XtremIO ;
- une tolérance aux pannes/protection élevée pour la haute disponibilité, systèmes en clusters tels qu' Oracle Real Application Clusters (RAC) et VMware vSphere High Availability (HA).

Introduction

Objectif

Ce livre blanc décrit une solution haute disponibilité évolutive pour Oracle Database 11g, dans le cadre du déploiement d'un environnement VMware vSphere virtualisé avec système de stockage XtremIO.

Périmètre

Les objectifs de ce livre blanc sont les suivants :

- présenter les principales technologies qui composent la solution ;
- décrire l'architecture et la conception de la solution ;
- en décrire et en valider les principaux composants et processus ;
- identifier ses principaux avantages métiers.

Public

Ce livre blanc s'adresse aux administrateurs BD Oracle, aux administrateurs VMware, aux administrateurs de stockage, aux architectes informatiques et aux responsables techniques chargés de la conception, de la création et de la gestion de l'infrastructure, des datacenters et des bases de données Oracle.

Présentation de la technologie

Présentation

Principaux composants technologiques de la solution :

- EMC XtremIO avec prise en charge de VMware VAAI ;
- VMware vSphere ;
- VMware vCenter ;
- EMC PowerPath®/VE ;
- Oracle Database 11gR2 Enterprise Edition ;
- Oracle RAC 11gR2.

EMC XtremIO

La baie de stockage EMC XtremIO est un système 100 % Flash qui repose sur une architecture scale-out. Il se compose de modules appelés X-Bricks, qui peuvent être assemblés afin d'accroître les performances et la capacité de la baie en fonction des besoins. Cette solution utilise deux modules X-Brick.

XtremIO représente une valeur ajoutée dans les principaux domaines suivants :

- **Performances** : quels que soient le taux d'occupation du système et l'utilisation de la capacité de stockage, la latence et le débit restent cohérents, prévisibles et constants. Au sein de la baie, la latence pour une demande d'E/S est généralement inférieure à une milliseconde. La Figure 1 présente un exemple de tableau de bord XtremIO utilisé pour surveiller les performances.

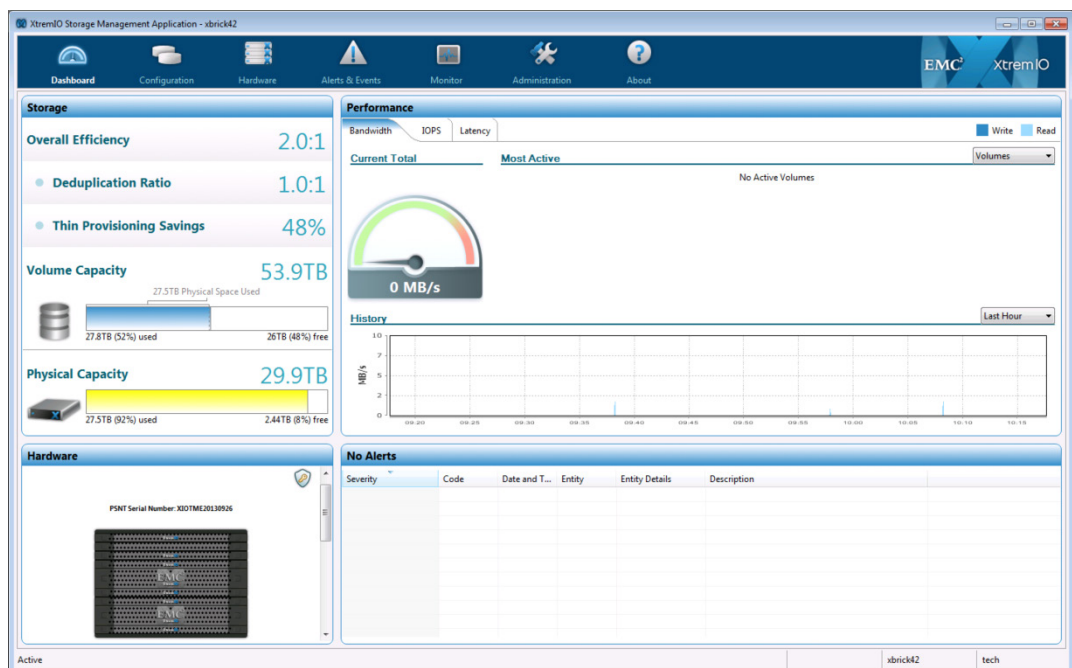


Figure 1. Tableau de bord de l'application de gestion du stockage XtremIO

- **Évolutivité** : le système de stockage XtremIO repose sur une architecture scale-out. Il commence par un module X-Brick unique. Si les performances et la capacité doivent être accrues, le système évolue grâce à l'ajout de modules X-Brick supplémentaires. Les performances progressent de manière linéaire : ainsi, deux modules X-Brick fournissent deux fois plus d'E/S par seconde que la configuration avec un seul X-Brick et quatre

modules X-Brick quatre fois plus. La latence reste faible à mesure que le système grandit. Les baies XtremIO effectuent un scale-out jusqu'à tous les niveaux requis de performances ou de capacité, comme le montre la Figure 2.

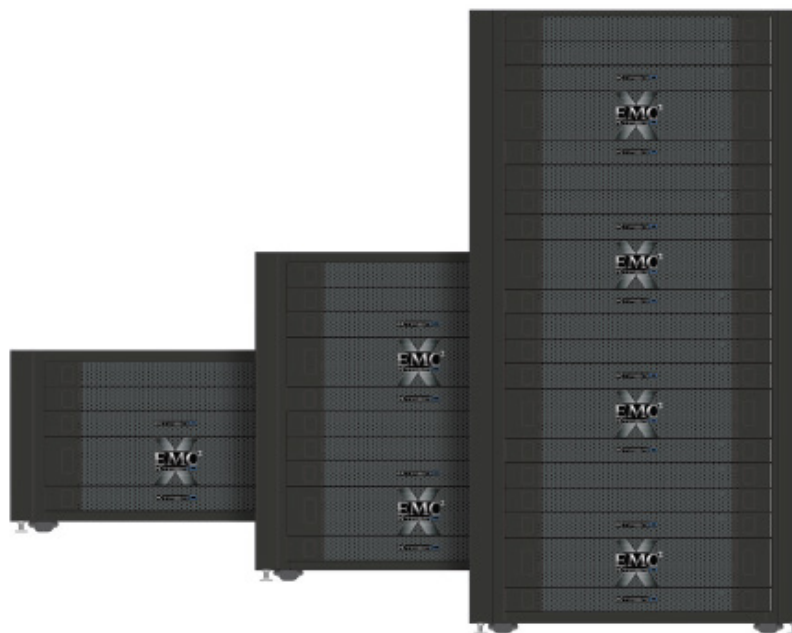


Figure 2. Unités de dimensionnement X-Brick

- **Protection des données :** XtremIO s'appuie sur un algorithme de protection des données exclusif et optimisé pour Flash, dénommé XDP (XtremIO Data Protection). Cet algorithme assure la protection des données tout en offrant des performances supérieures à tous les algorithmes RAID existants. L'optimisation de l'algorithme XDP permet de réduire le nombre d'écritures sur les médias Flash à des fins de protection des données.
- **Gestion des données :** XtremIO prend en charge les snapshots hautes performances avec optimisation de l'espace utilisé, la réduction des données à la volée, l'allocation dynamique et l'intégration complète de VMware vSphere VAAI avec prise en charge des protocoles Fibre Channel (FC) et iSCSI.
- **Simplicité :** pour le provisionnement du stockage, il suffit d'indiquer la taille de la LUN. Nul besoin de créer des groupes RAID, de choisir des disques SSD spécifiques sur lesquels créer les LUN, d'activer ou de désactiver l'allocation dynamique ou la déduplication. Pour plus d'informations, reportez-vous à la section Provisionnement des LUN.
- **Architecture équilibrée par nature :** XtremIO présente une conception scale-out. Par conséquent, il utilise plusieurs contrôleurs pour fournir des performances linéaires qui augmentent parallèlement à la capacité de la baie. L'un des objectifs essentiels de la conception d'XtremIO consiste à « maintenir une architecture naturellement équilibrée », afin que la baie puisse utiliser toutes les ressources de manière homogène, à tout moment. XtremIO permet d'atteindre cet objectif en équilibrant la charge applicative de la baie sur tous les contrôleurs, ce afin d'éviter les points sensibles et les processus de rééquilibrage back-end.

Voici quelques-unes des fonctions principales d'XtremIO :

Performances évolutives

XtremIO est conçu pour un déploiement scale-out afin de répondre aux besoins de performances et de capacité à venir, qu'ils découlent de nouvelles applications ou de celles déjà mises en œuvre. L'architecture XtremIO permet d'accroître les performances et la capacité par l'ajout de modules X-Brick, tout en conservant un seul point de gestion et en préservant l'équilibre des ressources sur l'ensemble du système.

Haute disponibilité

Éviter les pertes de données et maintenir le service en cas de défaillances multiples sont les principaux atouts de l'architecture XtremIO. Du point de vue matériel, aucun composant ne constitue un point unique de défaillance. Chaque module X-Brick est doté de deux alimentations. Le système possède également un double bloc-batteries de secours et deux ports réseau et de données. L'alimentation en entrée et les différents chemins de données font l'objet d'une surveillance constante, et toute défaillance déclenche une tentative de restauration ou une procédure de basculement sur incident.

L'architecture logicielle repose sur une construction similaire. Chaque élément d'information qui n'est pas associé à un disque SSD est conservé à plusieurs emplacements, dans un journal. Chaque module logiciel dispose de son propre journal, qui n'est pas conservé sur le même contrôleur de stockage, et peut servir à restaurer les données en cas de défaillance inattendue.

Réduction des données à la volée

XtremIO effectue une déduplication automatique et globale des données dès leur entrée dans le système. Cette technique permet de réduire la quantité de données écrites sur les médias Flash, ce qui allonge leur durée de vie et réduit les coûts. La réduction des données s'effectue en temps réel, et non en post-traitement. XtremIO ne fait intervenir aucun processus d'arrière-plan qui monopolise les ressources et n'effectue aucune opération de lecture/écriture supplémentaire (en relation avec la déduplication en post-traitement). Par conséquent, les performances sont améliorées, les ressources disponibles allouées aux E/S de l'hôte sont utilisées efficacement et la durée de vie des disques Flash est optimisée.

Snapshot XtremIO

Les snapshots XtremIO sont étroitement intégrés dans l'architecture de baie XtremIO. Leur implémentation repose entièrement sur les métadonnées et fait appel à la fonction de réduction des données pour empêcher leur copie au sein de la baie. Les snapshots XtremIO tirent parti des métadonnées en mémoire et des métadonnées à deux phases, avec un avantage supplémentaire : une structure de données efficace pour la prise en charge de la création rapide d'un snapshot fournissant les mêmes performances que les volumes de production, au niveau des lectures et des écritures, indépendamment de l'étendue ou de la profondeur de l'arborescence du snapshot.

Voici quelques exemples d'utilisation de base de données pouvant bénéficier des avantages des snapshots XtremIO :

- Sauvegarde de la base de données : en créant des snapshots des LUN de base de données et en les montant sur un serveur de sauvegarde dédié, les serveurs de production sont libérés des opérations de sauvegarde et peuvent faire une utilisation efficace des ressources au niveau des processus métiers.
- Protection contre la corruption des données logiques : en créant plusieurs snapshots sur de courts intervalles, les données logiques peuvent être récupérées à partir de ces snapshots. La création de snapshots n'implique pas une consommation supplémentaire de mémoire Flash ou de métadonnées, à moins que des données uniques non nulles ne soient écrites sur ces snapshots. Les utilisateurs peuvent ainsi créer des snapshots plus fréquemment et plus efficacement.
- Provisionnement rapide des environnements de test/dév. : la création et le montage du snapshot des LUN de base de données de production sur le serveur de test/dév. permettent le provisionnement rapide des copies exactes de la base de données de production à des fins de test/dév. ou autres.
- Déchargement du traitement : les opérations d'entrepôt décisionnel telles que l'extraction, la transformation et le chargement (ETL), ainsi que l'analytique en temps réel, peuvent tirer parti de la création et du montage des snapshots des LUN de base de données sur le serveur hors production. Ces snapshots permettent un accès rapide aux copies récentes des données de production.

Remarque : bien que cette solution soit conçue pour les bases de données Oracle, les snapshots sont tout aussi utiles pour les autres bases de données et applications.

Prise en charge de VMware VAAI

XtremIO s'intègre entièrement avec VMware vSphere via l'API vSphere VAAI pour le provisionnement et le clonage des machines virtuelles, le provisionnement des disques de machines virtuelles (VMDK) et le déploiement d'environnements de virtualisation à grande échelle. XtremIO propose des performances élevées, des temps de latence faibles et des délais de provisionnement réduits, quel que soit le provisionnement du stockage choisi au niveau VMDK. XtremIO prend en charge la primitive de remise à zéro des blocs VAAI et écrit les blocs de zéros de sorte à éliminer les problèmes de performances inhérents au provisionnement des volumes EZT (Eager-Zero-Thick) pour les disques virtuels. Outre la primitive de remise à zéro des blocs, XtremIO prend en charge d'autres primitives VAAI telles que XCOPY, ATS, UNMAP et l'allocation dynamique. Ainsi, le système XtremIO décharge les tâches de gestion du stockage VMware courante, pour une agilité du datacenter sans précédent :

- configuration et formatage rapide des volumes VMFS de toute taille ;
- utilisation de datastores volumineux pour une gestion simplifiée ;
- création rapide de datastores ;
- clonage instantané des machines virtuelles et vApps, avec déploiement en masse de centaines de nouvelles machines virtuelles en quelques minutes.

Seul XtremIO peut prendre en charge ces fonctionnalités. En effet, il propose une combinaison puissante de fonctionnalités : déduplication à la volée, métadonnées en mémoire et prise en charge VAAI.

- Pour une description plus détaillée des fonctions d’XtremIO, consultez le livre blanc *Introduction to the EMC XtremIO Storage Array*

VMware vSphere

La couche de virtualisation dissocie l’application des ressources physiques sous-jacentes. Elle offre une plus grande flexibilité dans la couche applicative en supprimant les périodes d’interruption matérielle pour maintenance, et permet de modifier le système physique sans nuire aux applications hébergées. Dans un exemple d’utilisation de virtualisation des serveurs, cette couche permet à plusieurs machines virtuelles indépendantes de partager le même matériel physique.

VMware vSphere constitue la plate-forme de virtualisation la plus complète et la plus robuste du secteur. Elle permet de virtualiser les applications critiques à l’aide de pools de ressources dynamiques, pour une flexibilité et une fiabilité sans précédent. Cette solution transforme les ressources physiques d’un ordinateur en virtualisant le CPU, la RAM, le disque dur et le contrôleur réseau. Cette transformation crée des machines virtuelles entièrement fonctionnelles qui exécutent des systèmes d’exploitation et des applications isolés et encapsulés.

VMware vCenter

VMware vCenter est un système de gestion centralisée pour l’infrastructure virtuelle VMware. Il vous fournit une interface unique accessible depuis différents périphériques à des fins de surveillance, de gestion et de maintenance de l’infrastructure virtuelle.

VMware vCenter gère également certaines fonctions avancées de l’infrastructure virtuelle VMware, comme VMware vSphere HA, VMware vSphere Distributed Resource Scheduling (DRS), VMware vMotion et VMware Update Manager.

EMC PowerPath/VE

EMC PowerPath/Virtual Edition (VE) for vSphere fournit des fonctions de multipathing optimisant les environnements virtuels VMware vSphere. PowerPath/VE est installé sur l’hôte VMware ESXi comme un module de noyau, et fonctionne comme un plug-in de multipathing (traitement massivement parallèle) qui fournit aux hôtes VMware ESXi des fonctions avancées de gestion des chemins et d’équilibrage de la charge.

Oracle Database 11g Enterprise Edition

Oracle Database 11g Enterprise Edition offre performances, évolutivité, sécurité et fiabilité pour une variété de serveurs uniques ou en cluster exécutant Windows, Linux ou UNIX. Cette solution prend en charge des fonctionnalités avancées, incluses ou disponibles en sus, qui ne sont pas disponibles avec Oracle Database 11gR2 Standard Edition. Il s’agit de fonctionnalités de sécurité telles que Virtual Private Database et d’options d’entrepôt décisionnel telles que le partitionnement et des fonctions d’analytique avancées.

Oracle RAC 11gR2

Dans Oracle RAC 11gR2, Oracle Grid Infrastructure combine Oracle Automatic Storage Management (ASM) et Oracle Clusterware en un ensemble de binaires distincts du logiciel de la base de données. Cette infrastructure offre à présent tous les services de cluster et de stockage requis pour exécuter une base de données Oracle RAC.

Oracle ASM

Oracle ASM est un système de fichiers de base de données et un gestionnaire de disques intégré compatible avec le système de cluster. Les fonctions de gestion de volumes et le système de fichiers ASM sont intégrés dans le noyau de la base de données Oracle. Dans Oracle RAC 11gR2, ASM a également été étendu pour inclure la prise en charge d'Oracle Cluster Registry (OCR) et des fichiers de vote à placer dans les groupes de disques ASM.

Oracle Clusterware

Oracle Clusterware est une solution de gestion de cluster portable intégrée dans la base de données Oracle. Cette solution offre l'infrastructure nécessaire pour exécuter Oracle RAC, y compris les services de gestion de cluster et les services haute disponibilité. Vous pouvez également l'utiliser pour faire bénéficier une application autre qu'Oracle de la haute disponibilité dans le cluster.

Architecture de la solution

Présentation

Cette section décrit l'architecture de la solution.

Celle-ci représente un rapport coûts/performances optimal dans les environnements applicatifs critiques. La base de données Oracle est déployée sous forme de base de données RAC virtualisée à quatre nœuds, sur une baie de stockage XtremIO composée de deux volumes XtremIO X-Brick.

Schéma de l'architecture

L'organisation de la solution est illustrée dans la Figure 3.

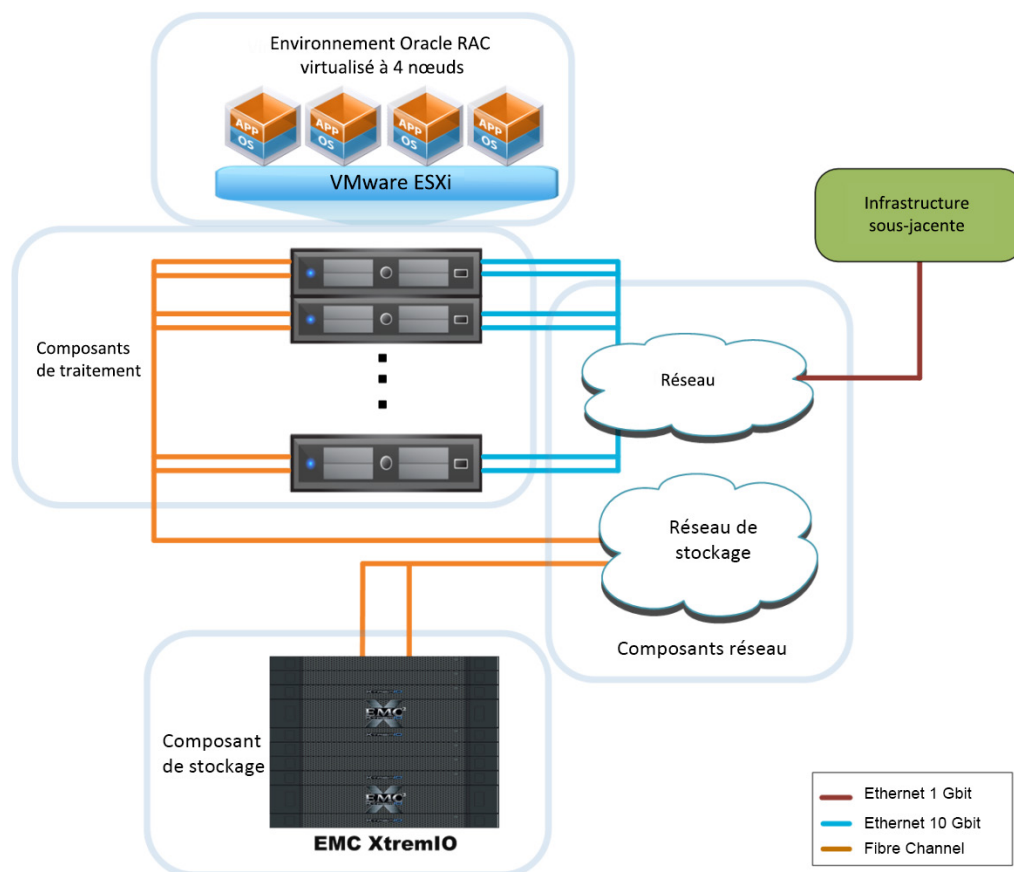


Figure 3. Architecture de la solution

La solution se compose des éléments suivants :

- **Couche de traitement** : comprend quatre serveurs utilisant un total de 32 cœurs, avec des processeurs cadencés à 2,9 GHz et 2 To de RAM. Ces serveurs montés en rack permettent d'adopter une approche orientée hautes performances, consolidation et virtualisation des infrastructures Oracle, ce qui améliore la flexibilité du déploiement sans devoir modifier les applications.
- **Couche réseau** : comprend deux switches IP et deux switches SAN de classe directeur, que nous avons¹ configurés pour obtenir une bande passante active de 108 Go/s. Les switches SAN sont conçus pour un déploiement dans des réseaux de stockage prenant en charge les Clouds d'entreprise et les datacenters virtualisés.

¹ Dans ce livre blanc, « nous » désigne l'équipe d'ingénieurs EMC Solutions qui a validé la solution.

- **Couche de stockage** : comprend deux baies X-Brick (12U) avec 14,94 To de capacité physique utile.

Les serveurs sont installés avec VMware vSphere et configurés en tant que cluster VMware ESXi.

Quatre machines virtuelles sont créées sur ce cluster VMware ESXi, qui exécute une base de données Oracle RAC à quatre nœuds. Nous avons examiné les performances du système en exécutant séparément les charges applicatives OLTP et d'entrepôt décisionnel (DW) sur la base de données Oracle RAC.

Ressources matérielles

Le Tableau 1 répertorie les ressources matérielles de la solution.

Tableau 1. Ressources matérielles

Matériel	Quantité	Configuration
Baie de stockage	1	Baie XtremIO composée de deux modules X-Brick
Serveurs	4	8 cœurs, processeurs 2,9 GHz, 512 Go de RAM, avec : <ul style="list-style-type: none"> • 1 carte réseau 1 Gigabit Ethernet (GbE) • 1 carte réseau 10 GbE
Switches LAN	2	10 GbE
Switches SAN	2	FC

Ressources logicielles

Le Tableau 2 répertorie les logiciels de la solution.

Tableau 2. Ressources logicielles

Logiciel	Version	Remarques
VMware vSphere	5.0	Hypervisor hébergeant toutes les machines virtuelles
VMware vCenter	5.1	Gestionnaire vSphere
Red Hat Enterprise Linux	6.3	Système d'exploitation des serveurs de base de données
Oracle Database 11g Release 2	Enterprise Edition 11.2.0.3	Base de données
Oracle Grid Infrastructure 11g Release 2	Enterprise Edition 11.2.0.3	Clusterware avec ASM pour la gestion des volumes
Silly Little Oracle Benchmark (SLOB)	2	Outil de benchmark OLTP
XIOS	2.2 SP3	Système d'exploitation XtremIO

Couche de stockage : EMC XtremIO

Conception de cluster

EMC XtremIO s'appuie sur une conception de cluster scale-out optimisée pour le Flash qui ajoute de manière linéaire de la capacité et des performances pour répondre aux besoins en matière de stockage. Chaque module X-Brick est doté de contrôleurs de stockage haute disponibilité, avec configuration entièrement actif/actif, sans aucun point unique de défaillance. Des modules X-Brick supplémentaires peuvent être ajoutés à un système existant et regroupés sur un backbone de réseau redondant, haute disponibilité et présentant un temps de latence extrêmement faible. Dans un tel système, les performances évoluent de manière linéaire, alors que la latence maintient un niveau bas.

Le système d'exploitation XtremIO (XIOS) gère les clusters de stockage et fournit les fonctionnalités suivantes :

- Il garantit une répartition équitable de la charge sur tous les disques SSD du cluster afin d'obtenir les meilleurs niveaux de performance et d'endurance possible pour répondre aux charges applicatives exigeantes pendant toute la durée de vie de la baie.
- Il rend inutile la configuration complexe habituelle des baies traditionnelles. Il élimine le besoin de définir les niveaux RAID, la taille des groupes de disques, la largeur des bandes, les règles de mise en cache, et de créer des agrégats ou de procéder à d'autres tâches de configuration manuelles.
- Il configure les volumes de manière automatique et optimale et garantit l'augmentation automatique des performances des E/S sur les volumes existants et sur les Datasets en cas de scale-out du cluster.

Réduction des données à la volée

La baie XtremIO 100 % Flash exécute la déduplication des données à la volée en fonction d'un algorithme qui vérifie que chaque bloc de données stocké sur la baie n'est pas identique à du contenu existant. Ainsi, chaque E/S de stockage est dédupliquée en temps réel à la réception et seuls les blocs uniques sont écrits sur le stockage Flash. En outre, la déduplication sur la baie XtremIO améliore les performances, car les cycles d'E/S des disques SSD ne sont jamais consommés par l'écriture de blocs en double ou le traitement de données non dédupliquées. Cela préserve la capacité d'E/S maximale pour la gestion des demandes provenant des hôtes.

Lorsque le mode archivelog est activé et qu'un journal redo log en ligne est plein, le processus d'archivage copie les données redo à partir de ce redo log en ligne vers le fichier log d'archivage. Le redo log plein n'est plus disponible en écriture pour le processus Log Writer jusqu'à son archivage. Le log d'archivage est une copie exacte du redo log en ligne. Par conséquent, la baie n'exécute pas d'E/S d'écritures physiques sur les disques SSD back-end ; elle envoie les mises à jour vers des pointeurs in-memory et l'opération d'écriture se termine instantanément.

Une fois que le Log Writer (LGWR) a écrit les nouvelles données dans le journal redo log en ligne en cours d'archivage, le pointeur en mémoire du fichier redo log en ligne est mis à jour et pointe vers les blocs redo physiques nouvellement alloués, qui stockent les dernières données redo. Le pointeur en mémoire du journal redo log archivé est intact, ce qui signifie que les blocs physiques utilisés par le journal redo log archivé ne sont pas obsolètes tant qu'ils existent dans le log d'archivage de destination.

XtremIO permet également une mise en cache compatible avec la déduplication : les blocs conservés dans le cache peuvent être utilisés pour n'importe quelle référence logique à ces blocs. La mise en cache compatible avec la déduplication, alliée à la déduplication à la volée, réduit considérablement les temps de latence associés à la gestion de situations complexes, telles que les démarrages simultanés de plusieurs machines virtuelles, en permettant un accès aux données cohérent, en moins d'une milliseconde.

Allocation dynamique

En plus de proposer des performances élevées, XtremIO offre des fonctions d'allocation dynamique qui permettent d'allouer la capacité à la demande, en fonction des besoins des applications, sans opération de post-récupération ou d'incidence sur les performances des E/S de stockage sur la baie. L'allocation dynamique d'XtremIO est également granulaire : la capacité est allouée en blocs afin de garantir une utilisation parcimonieuse du stockage Flash, ce qui correspond à l'utilisation des tailles de blocs d'E/S par vSphere.

Protection contre les pannes

Les composants totalement redondants d'XtremIO offrent fiabilité et disponibilité et permettent de réagir à toutes les pannes matérielles dans perte de service. XtremIO inclut les fonctions de protection contre les pannes suivantes :

- double alimentation dans les contrôleurs et les boîtiers DAE pour gérer la perte d'une alimentation tout en maintenant le contrôleur/boîtier DAE en service ;
- contrôleurs redondants à configuration actif/actif pour la prise en charge des pannes de contrôleur ;
- modules d'interconnexion SAS (Serial-Attached SCSI) redondants dans les boîtiers DAE ;
- liaisons de communication intercontrôleurs redondantes ;
- plusieurs connexions hôtes avec fonctions de multipathing pour gérer les défaillances de chemins ;
- XDP pour gérer les pannes de plusieurs disques SSD ;
- plusieurs techniques permettant de garantir une intégrité des données constante.

Cette conception tolérante aux pannes convient parfaitement au déploiement de systèmes de clusters haute disponibilité tels qu'Oracle RAC 11g.

Évolutivité

Les clusters de stockage XtremIO prennent en charge une conception scale-out entièrement distribuée qui permet l'augmentation linéaire de la capacité et des performances pour une infrastructure agile. XtremIO a recours à une approche modulaire dans laquelle la baie évolue via l'ajout de modules X-Brick supplémentaires.

XtremIO fournit un accès aux hôtes via des contrôleurs N-way actifs/actifs pour une évolutivité linéaire des performances et de la capacité, à des fins de prise en charge simplifiée des environnements virtualisés en pleine expansion. Par conséquent, à mesure que la capacité croît dans la baie, il en va de même pour les performances via l'ajout de contrôleurs de stockage avec une latence cohérente.

Opérations sur les métadonnées in-memory

Le cluster XtremIO distribue les métadonnées équitablement entre tous les contrôleurs de stockage, les maintenant en mémoire pendant l'exécution. Elles sont envoyées vers les disques SSD pour permettre à la baie de tolérer les pannes et les coupures d'alimentation, mais pendant les heures de fonctionnement normales, toutes les recherches de métadonnées sont basées sur la mémoire. Cela n'est rendu possible que par la segmentation des tables de métadonnées et leur répartition homogène sur tous les contrôleurs de stockage. Par opposition, une conception à deux contrôleurs pourrait ne pas contenir suffisamment de RAM pour stocker toutes les métadonnées en mémoire et elle requerrait le reclassement de grandes quantités de métadonnées vers la mémoire Flash, ce qui nuirait aux performances.

Les métadonnées in-memory et le modèle de déduplication à la volée unique d'XtremIO s'allient pour offrir des fonctionnalités inédites dans les datacenters virtualisés.

XtremIO Management Server

Avec XtremIO Management Server (XMS), vous pouvez contrôler et gérer le système. Fonctions offertes par XMS :

- initialisation et formatage de nouveaux systèmes ;
- surveillance de l'intégrité du système et des événements ;
- surveillance des performances système ;
- maintenance d'une base de données historique avec statistiques de performances ;
- fourniture de services GUI et CLI aux clients ;
- implémentation d'opérations de gestion des volumes et de protection des données ;
- maintien en conditions opérationnelles (arrêt, démarrage et redémarrage) du système.

XMS est préinstallé et accessible via une interface utilisateur simple d'utilisation, XtremIO Storage Management Application (XSMA), et il se connecte aux ports de gestion des contrôleurs de stockage X-Brick via TCP/IP. Le cluster XtremIO gère le stockage indépendamment de la déconnexion d'un serveur XMS.

XMS est déployé en tant que machine virtuelle dans l'environnement de gestion du système, qui s'appuie sur un cluster VMware HA. Ainsi, la machine virtuelle XMS peut rapidement redémarrer en cas de panne. Si votre environnement n'est pas virtualisé, un serveur physique peut tenir lieu de XMS.

Configuration du stockage

Dans le cadre de cette solution, la baie XtremIO est déployée avec deux modules X-Bricks, interconnectés via des switches QDR InfiniBand 40 Gbit/s redondants intégrés qui fournissent une connectivité back-end entre les contrôleurs de stockage. Cela garantit un réseau haute disponibilité, avec des temps de latence très faibles.

La baie XtremIO de ce test contient cinquante disques SSD de 400 Go. Ils sont configurés avec XDP pour atteindre une capacité physique de 14,94 To, comme le montre la Figure 4.

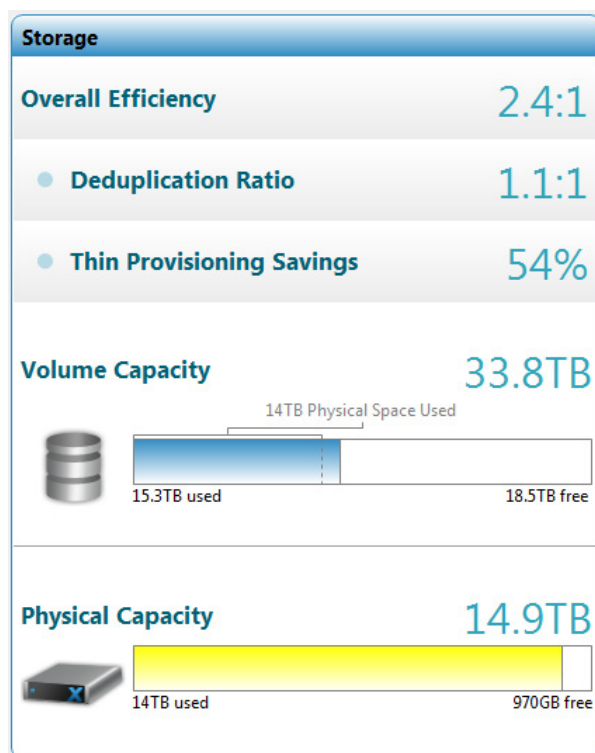


Figure 4. Interface utilisateur XtremIO : panneau de stockage

Conception du stockage

Avec la conception du stockage traditionnelle des bases de données Oracle, plusieurs groupes RAID contenant différents types de disque sont créés, chacun avec divers niveaux de protection et distribué sur plusieurs contrôleurs.

Avec XtremIO, tous les disques sont protégés par XDP, et les données de la baie sont automatiquement distribuées sur l'ensemble des modules X-Brick afin de préserver la prévisibilité des performances et un niveau d'usure équivalent sur tous les supports Flash.

Habituellement, les bases de données génèrent deux types d'E/S : E/S aléatoires et E/S séquentielles. Avec XtremIO, ces E/S sont traitées de la même manière, car les données sont randomisées et distribuées de façon homogène dans l'ensemble de la baie.

Dans le cadre de cette solution, la taille de volume est standardisée en fonction de l'utilisation des données, comme le montre le Tableau 3.

Tableau 3. Conception du stockage Oracle sur les modules X-Brick XtremIO

Fonction du volume	Taille (Go)
Disques Oracle ASM pour les fichiers de données	700
Disques Oracle ASM pour les journaux redo log	40
Disques Oracle ASM pour les fichiers Cluster Ready Service (Oracle Cluster Registry et fichiers de disque de vote)	10
Disques Oracle ASM pour Oracle Automatic Storage Management Cluster File System (ACFS) pour le chargement des données de fichiers	200

Pour la base de données OLTP, les volumes ont été créés et présentés aux serveurs VMware ESXi pour une utilisation avec Oracle Database 11g, comme le montre le Tableau 4.

Tableau 4. Attribution des volumes/LUN pour la base de données OLTP

Nom du dossier	Taille du volume (Go)	Nombre de volumes	Groupe de disques ASM	Redondance du groupe de disques ASM	Nombre de groupes d'initiateurs
Data	700	8	+DATA	Externe	4
Log	40	4	+REDO	Externe	4
CRS	10	2	+CRS	Externe	4

Provisionnement des LUN

Le provisionnement des LUN dans XtremIO est simple et direct. Vous pouvez créer des LUN et les présenter aux serveurs en quelques clics à partir de la console de gestion de la baie XtremIO. Les étapes suivantes décrivent comment créer des LUN et les mapper aux initiateurs.

1. Dans la console de gestion de la baie XtremIO, sélectionnez l'élément Configuration, cliquez sur **Add**, spécifiez le nom et la taille des LUN dans la fenêtre contextuelle et cliquez sur **Next**, comme le montre la Figure 5.

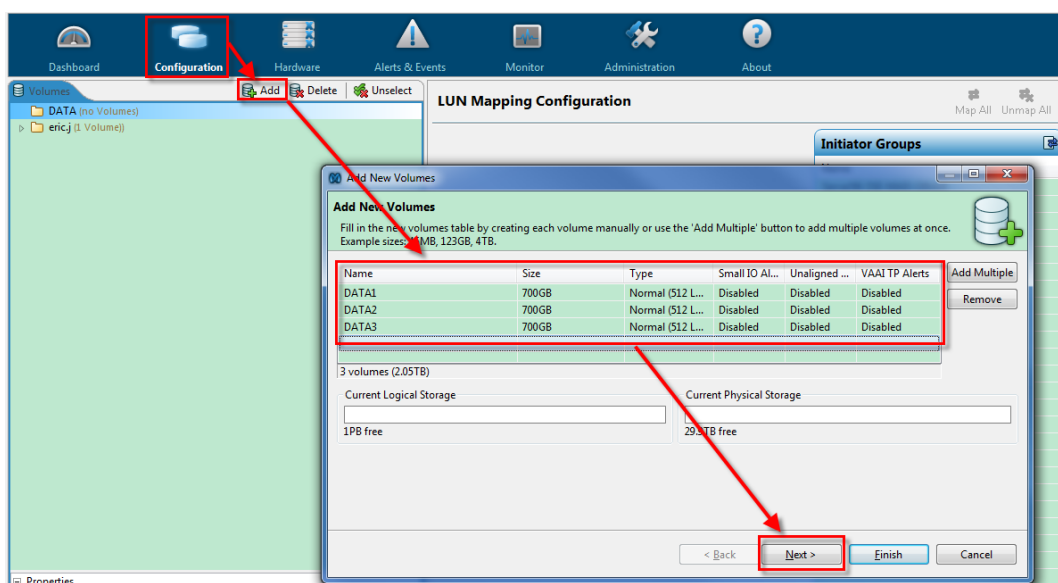


Figure 5. Initier la création des LUN et spécifier leur nom et leur taille

2. Sélectionnez le dossier parent des LUN et cliquez sur **Finish**, comme le montre la Figure 6.

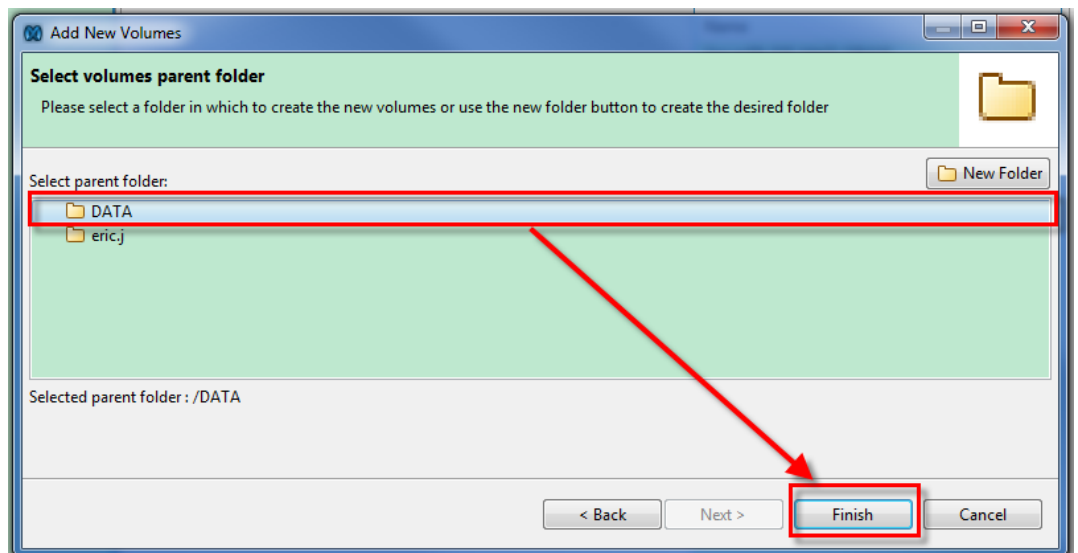


Figure 6. Spécifier le dossier des LUN

3. Sélectionnez les LUN nouvellement créées et les Initiator Groups ou groupes d'initiateurs, puis cliquez sur **Map All**, comme le montre la Figure 7.

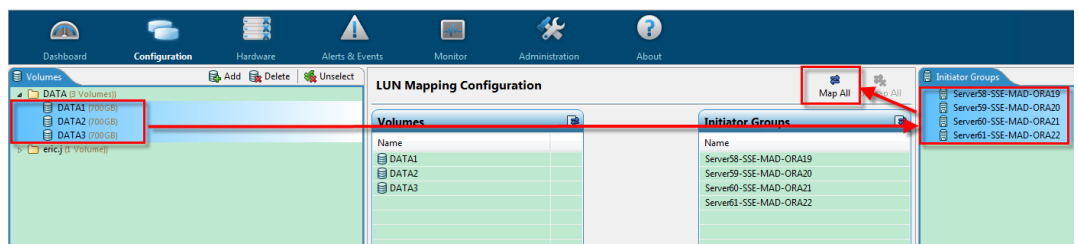


Figure 7. Mapper les LUN aux groupes d'initiateurs

4. Cliquez sur **Apply**. Toutes les LUN créées sont mappées sur les groupes d'initiateurs, comme le montre la Figure 8.

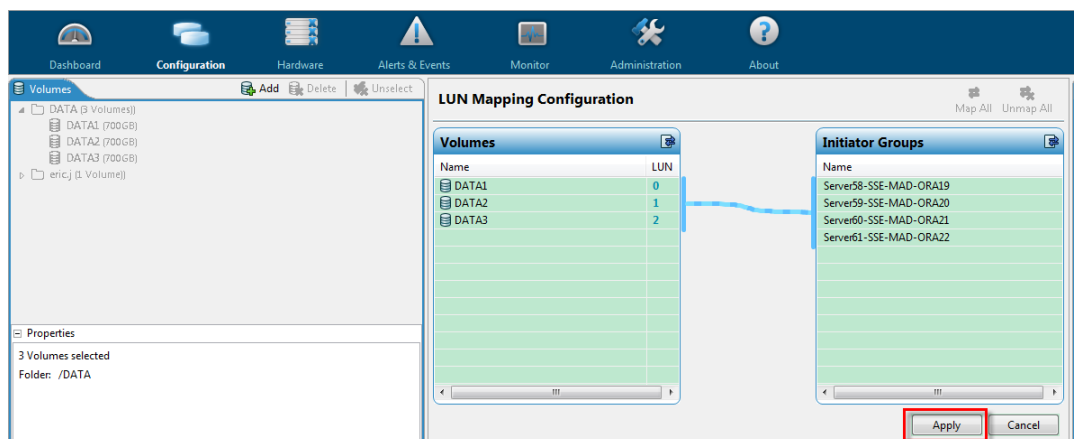


Figure 8. Appliquer la configuration du mappage des LUN

5. Après le mappage des LUN nouvellement créées sur les initiateurs, réanalysez les adaptateurs de stockage sur les hôtes VMware ESXi afin que les LUN soient visibles.

La Figure 9 présente les volumes Data11 à Data18, regroupés sous le dossier Data. Ces volumes comprennent le groupe de disques ASM +DATA de la base de données OLTP. Dans le cadre de cette solution, un groupe d'initiateurs se compose d'un serveur physique unique, ainsi que de ses ports FC. Le volume Data11 est présenté en tant que LUN et mappé aux quatre groupes d'initiateurs.

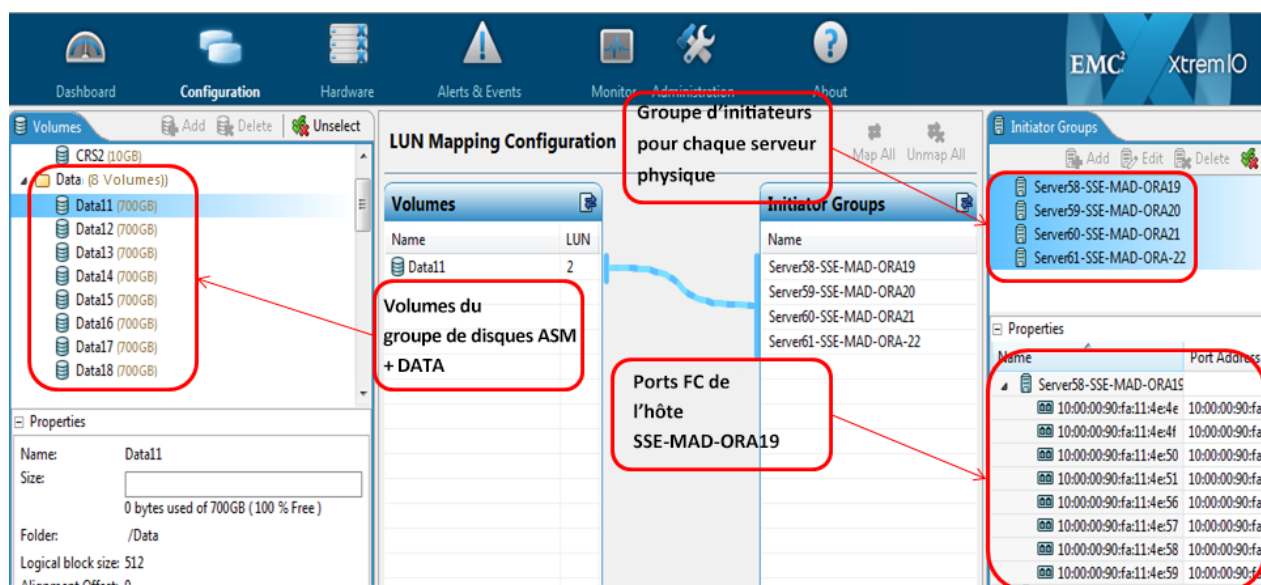


Figure 9. Interface utilisateur XtremIO - Mappage des volumes/LUN sur les serveurs

Pour la base de données d'entrepôt décisionnel, les volumes sont créés et groupés dans des dossiers, comme le montre le Tableau 5. Chaque volume est présenté en tant que LUN aux groupes d'initiateurs, comme le montre la Figure 9.

Tableau 5. Conception des groupes de disques ASM pour la base de données d'entrepôt décisionnel

Nom du dossier	Taille du volume (Go)	Nombre de volumes	Groupe de disques ASM	Redondance du groupe de disques ASM	Nombre de groupes d'initiateurs
DWDATA	700	8	+DATA	Externe	4
DWLog	40	4	+REDO	Externe	4
DWCRS	10	2	+CRS	Externe	4
DWACFS	200	1	+CSV	Externe	4

Couche réseau

Présentation

Cette section décrit brièvement les bonnes pratiques suivies dans cette solution pour la configuration des réseaux SAN et IP, et pour la configuration réseau d'un serveur VMware ESXi. Lorsque vous déployez une solution en cluster virtualisée, comme Oracle RAC, EMC vous recommande de prendre en compte la redondance de traitement et la redondance réseau lors de la conception de votre réseau pour la tolérance aux pannes.

Bonnes pratiques SAN

EMC vous recommande de mettre en œuvre les bonnes pratiques suivantes en matière de SAN :

- Utilisez plusieurs adaptateurs HBA sur les serveurs VMware ESXi et deux switches SAN pour fournir plusieurs chemins entre les hôtes et le cluster XtremIO.
- Zonez chaque port FC des serveurs de base de données vers les ports des modules X-Brick XtremIO pour obtenir haute disponibilité et performances élevées.
- Utilisez un logiciel de gestion des chemins et de multipathing dynamique (tel que PowerPath/VE) sur les hôtes pour permettre au processus de basculement sur incident d'accéder à des chemins d'accès de substitution et pour assurer un équilibrage de la charge.
- Si ce type de logiciel n'est pas utilisé, nous vous recommandons de configurer un multipathing actif/actif avec permutation circulaire pour la règle de sélection des chemins sur les hôtes VMware ESXi.

Bonnes pratiques relatives au réseau IP

EMC vous recommande de mettre en œuvre les bonnes pratiques suivantes en matière de réseau IP :

- Utilisez plusieurs cartes réseau et switches pour assurer la redondance réseau.
- Utilisez une connexion réseau 10 GbE, le cas échéant.
- Utilisez des LAN virtuels pour regrouper logiquement les périphériques qui se trouvent sur différents segments réseau ou sur des sous-réseaux.
- Activez et configurez les trames Jumbo² sur les piles physiques et virtuelles pour les réseaux 10 GbE.

Bonnes pratiques relatives au réseau VMware vSphere

La gestion réseau dans les environnements virtuels et physiques implique les mêmes concepts et nombre de bonnes pratiques s'appliquent à ces deux types d'environnement. Cependant, d'autres considérations sont applicables en termes de segmentation du trafic, de disponibilité et de débit en cas de virtualisation.

Cette solution a été conçue pour gérer efficacement plusieurs réseaux ainsi que la redondance des adaptateurs réseau sur les hôtes VMware ESXi. Les principales recommandations liées aux bonnes pratiques sont les suivantes :

- Isolez le trafic de l'infrastructure de celui des machines virtuelles à des fins de sécurité.

² Les MTU d'une taille supérieure à 1 500 octets sont qualifiées de « trames Jumbo ». Les trames Jumbo requièrent une connexion Gigabit Ethernet sur la totalité de l'infrastructure réseau, y compris au niveau des serveurs, des switches et des serveurs de base de données.

- Utilisez la famille d'adaptateurs réseau virtuels VMXNET3.
- Regroupez les cartes réseau physiques pour de meilleures performances et redondance réseau. Par exemple, utilisez une paire de cartes réseau physiques par serveur/vSwitch et liez chacune d'elles via uplink à des switches physiques distincts.

Pour plus d'informations sur la mise en réseau avec VMware vSphere, suivez les instructions du document *VMware vSphere Networking*.

Couche de traitement

Présentation

Le choix d'une plate-forme de serveur pour une infrastructure virtualisée se base à la fois sur les capacités de prise en charge de la plate-forme et sur les exigences techniques de l'environnement. Dans les environnements de production, les serveurs doivent présenter les caractéristiques suivantes :

- cœurs de processeur et mémoire suffisants pour prendre en charge le nombre et la charge applicative requis des machines virtuelles ;
- une connectivité suffisante, Ethernet et FC, pour permettre une connectivité redondante aux switches de réseau IP et de stockage ;
- une capacité suffisante pour supporter une panne de serveur et prendre en charge le basculement sur incident des machines virtuelles.

Dans cet environnement de test, nous avons utilisé quatre serveurs physiques configurés dans un cluster VMware vSphere HA, chacun exécutant un serveur VMware vSphere ESXi. Nous avons ensuite configuré quatre machines virtuelles pour concevoir un déploiement de base de données Oracle RAC virtualisée à quatre nœuds.

Oracle Database 11gR2 et les versions supérieures sont entièrement prises en charge en cas de déploiement sur la technologie VMware vSphere ESXi³. Cela inclut le déploiement d'une configuration Oracle RAC.

Ressources de traitement et de stockage

EMC vous recommande d'implémenter les bonnes pratiques suivantes au niveau des ressources de traitement VMware, comme décrit dans le document *Oracle Databases on VMware - Best Practices Guide* :

- Utilisez la technologie NUMA (Non-Uniform Memory Access) sur les serveurs VMware ESXi, une architecture informatique dans laquelle l'accès à une mémoire plus proche d'un processeur particulier est plus rapide que l'accès à une mémoire plus éloignée.
- Allouez une quantité de mémoire vRAM aux machines virtuelles inférieure ou égale à la mémoire locale du nœud NUMA (processeur).
- Installez VMware Tools, qui comprend plusieurs utilitaires améliorant les performances du système d'exploitation invité de la machine virtuelle et la facilité de gestion de cette dernière.
- Allouez une vRAM au moins deux fois supérieure à la zone SGA (System Global Area) Oracle.
- Configurez les réserves de mémoire des machines virtuelles de sorte qu'elles atteignent au minimum la taille de la zone SGA Oracle et les frais supplémentaires associés au système d'exploitation.
- Utilisez plusieurs contrôleurs paravirtualisés SCSI (PVSCSI) pour les machines virtuelles de base de données. L'utilisation de plusieurs contrôleurs SCSI virtuels permet l'exécution de différentes opérations d'E/S parallèles au sein du système d'exploitation invité. Configurez le contrôleur PVSCSI pour les disques utilisés pour stocker les fichiers de données de base de données.

³ Le document My Oracle Support Doc ID 249212.1 définit la politique de support VMware d'Oracle.

- Isolez le trafic des E/S des journaux redo log Oracle de celui des E/S des fichiers de données à l'aide de contrôleurs SCSI virtuels distincts. La bonne pratique consiste à utiliser un contrôleur pour le système d'exploitation et les LUN de fichiers d'échange, un autre contrôleur pour les LUN de journaux redo log des bases de données, et un ou plusieurs contrôleurs supplémentaires pour les LUN de fichiers de données des bases de données.
- Configurez les Huge Pages dans le système d'exploitation invité. L'utilisation des Huge Pages entraîne des frais supplémentaires réduits au niveau de la gestion de la mémoire et des performances d'Hypervisor améliorées.
- Augmentez la longueur de la file d'attente de l'adaptateur HBA et du contrôleur SCSI pour optimiser le nombre d'E/S parallèles vers la baie de stockage. Consultez l'*article de la base de connaissances VMware 2053145* pour en savoir plus sur le réglage de la longueur de la file d'attente.

Virtualisation

Virtualisation du réseau

Sur chaque serveur VMware ESXi, nous avons créé deux vSwitches standard avec une configuration commune, comme le montre le Tableau 6.

Tableau 6. Configuration du vSwitch

Nom	Objectif
vSwitch0	Trafic de gestion et des machines virtuelles publiques
vSwitch1	Configuration tolérante aux pannes pour l'interconnexion Oracle RAC

Nous avons utilisé le pilote VMXNET3 hautes performances pour attribuer deux cartes réseau (1 GbE et 10 GbE) à chaque machine virtuelle. Nous avons mappé la carte réseau virtuelle 1 GbE sur le vSwitch0 pour fournir du trafic public, et la carte réseau virtuelle 10 GbE sur le vSwitch1 pour fournir le trafic d'interconnexion Oracle RAC.

Configuration des modèles de machine virtuelle

Les modèles VMware réduisent le délai du déploiement et automatisent les tâches d'installation et de configuration répétitives sur chaque machine virtuelle requise. Grâce à ce modèle, vous pouvez déployer une machine virtuelle contenant des utilisateurs d'applications et un système d'exploitation préconfigurés, ainsi qu'un logiciel configuré et prêt à l'emploi avec intervention minimale de l'utilisateur.

Les spécifications de personnalisation stockées dans VMware vCenter simplifient le déploiement des machines virtuelles. Un assistant de déploiement, un outil d'automatisation ou un script peut utiliser ces modèles pour créer ou modifier automatiquement des paramètres serveur (tels que le nom du serveur, le fuseau horaire et la configuration réseau) avant de générer la nouvelle machine virtuelle.

Configurez le modèle de machine virtuelle dans le client VMware vSphere conformément aux exigences et conditions requises au niveau des logiciels Oracle, décrites dans le Tableau 7.

Tableau 7. Configuration des modèles de machine virtuelle

Composant	Description
CPU	8 vCPU pour la charge applicative OLTP
Mémoire	128 Go
Système d'exploitation	Red Hat Enterprise Linux Server version 6.3
Noyau	2.6.32-279.el6
Interfaces réseau virtuelles	Eth0 : réseau IP public/de gestion Eth1 : dédié à l'interconnexion de clusters
Utilisateur OS (utilisateur créé et mot de passe défini)	Nom d'utilisateur : oracle ID utilisateur : 1101
Groupes OS	Groupe : oinstall ID groupe : 1000 Groupe : dba ID groupe : 1031
Logiciel préinstallé	Base de données Oracle avec Oracle Grid Infrastructure (pour serveur autonome)
Packages RPM installés (conditions requises Oracle)	Comme indiqué dans le guide d'installation Oracle
Configuration des disques	Disque virtuel de 128 Go pour les dossiers racine, /tmp, l'espace de fichier d'échange et les fichiers binaires Oracle 11gR2 Database

Après le déploiement des machines virtuelles, nous avons ajouté le disque dédié au stockage des bases de données en tant que disque RDM (Raw Device Mapping).

Autorisation d'accès aux périphériques de stockage

Sur chaque machine virtuelle, nous avons ajouté les LUN de stockage de la base de données de la baie XtremIO en tant que périphériques RDM et nous les avons répartis sur quatre contrôleurs PVSCSI pour équilibrer les E/S, comme le montre la Figure 10.

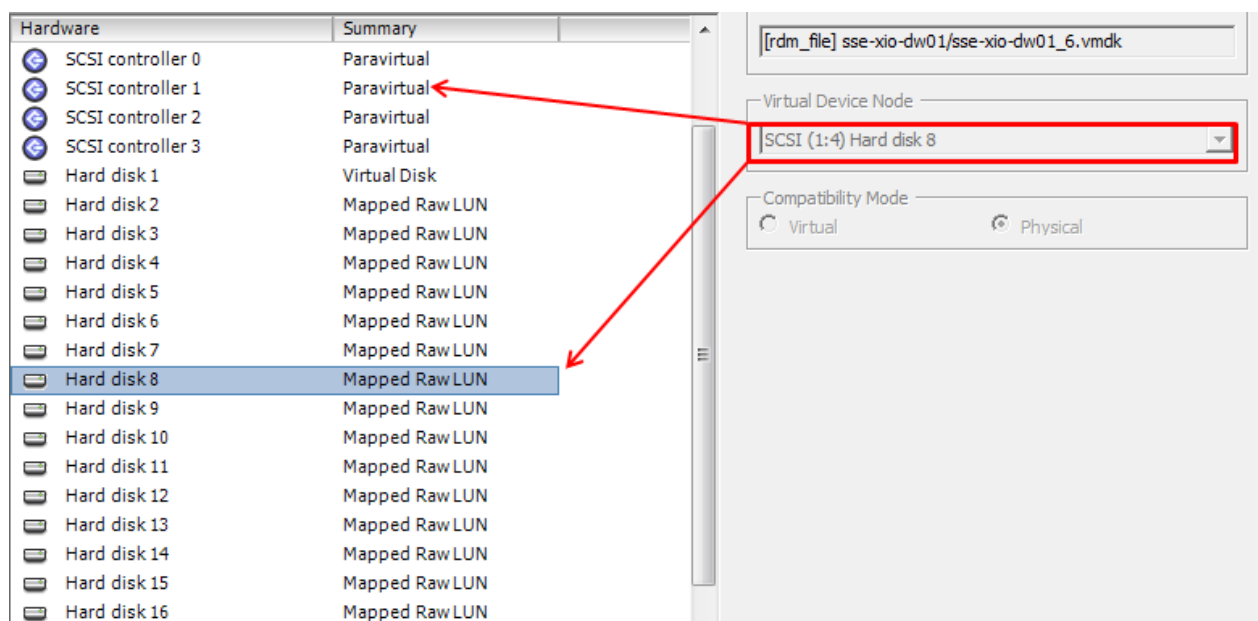


Figure 10. Configuration des disques dans la machine virtuelle

Nous avons défini l'option **SCSI Bus Sharing** des contrôleurs SCSI sur **Physical** afin que les machines virtuelles bénéficient d'un accès direct et partagé aux LUN RDM sur les serveurs VMware ESXi, comme le montre la Figure 11.

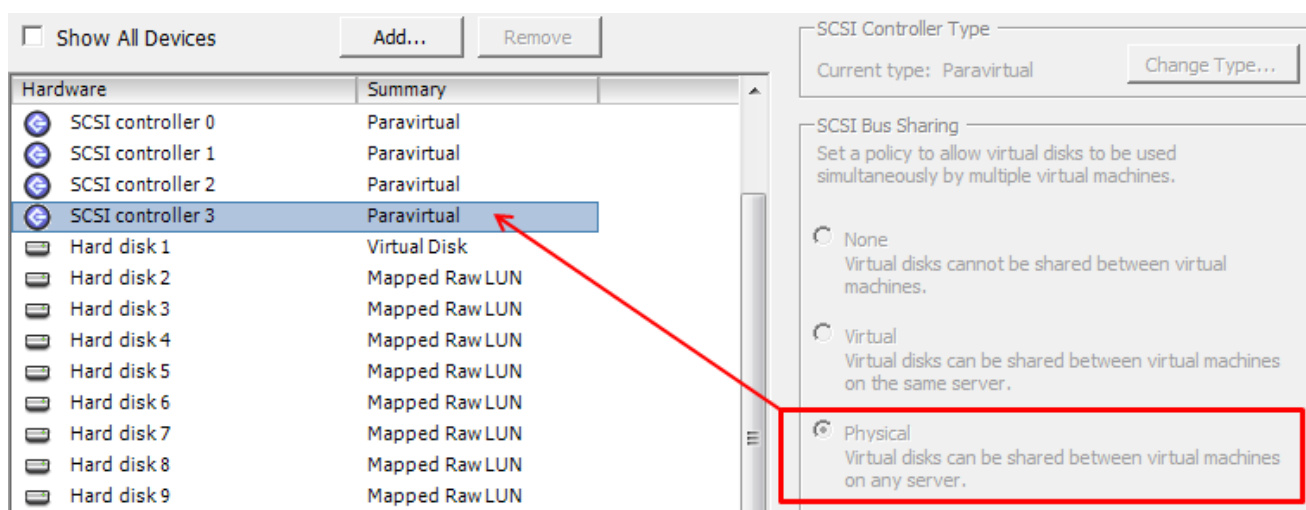


Figure 11. Paramètres SCSI Bus Sharing

Base de données Oracle

Présentation

Dans la cadre de cette solution, nous avons créé une base de données OLTP Oracle RAC à quatre nœuds virtualisée et une base de données d'entrepôt décisionnel, toutes deux à l'aide d'Oracle ASM et d'Oracle ACFS, sur VMware vSphere.

Dans Oracle Database 11g, Oracle ASM et Oracle Clusterware sont intégrés avec l'infrastructure Oracle Grid Infrastructure. ACFS étend les fonctionnalités d'ASM en fournissant un système de fichiers de consigne basé sur extent, robuste et générique. Dans cette solution, nous avons utilisé ASM pour stocker les fichiers de base de données et ACFS pour stocker les fichiers CSV pour le test de chargement des données.

La Figure 12 présente les instances ASM de l'environnement OLTP Oracle RAC.

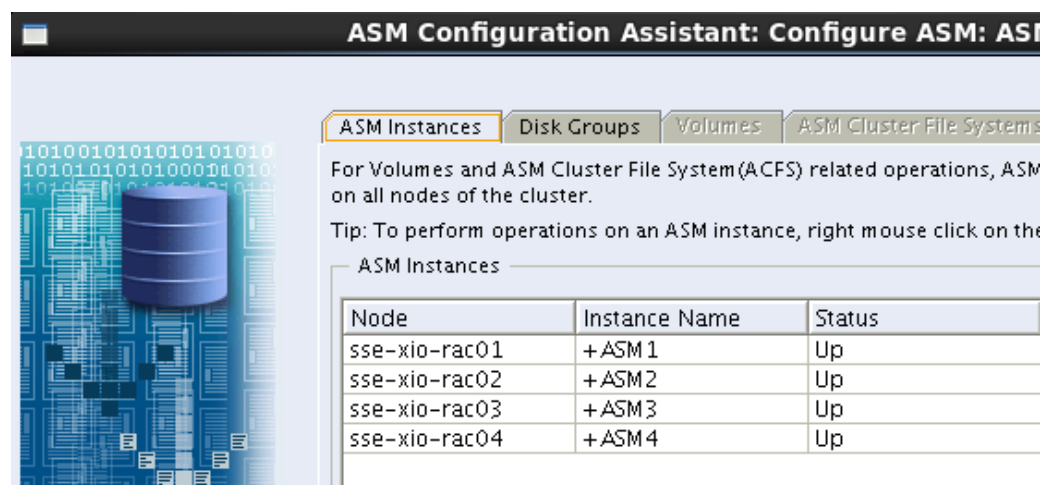


Figure 12. Instances Oracle ASM

Remarque : la base de données utilisée pour le test de l'environnement OLTP ou DW a été configurée en désactivant la consigne des archives pour atteindre des performances optimales.

Conception du stockage des bases de données OLTP

Comme le montre le [Tableau 4 page 21](#), nous avons utilisé trois groupes de disques ASM pour stocker les fichiers de base de données pertinents, y compris les fichiers de données, fichiers de contrôle, journaux redo log en ligne, fichiers de disque de vote, fichiers CRS et fichiers temporaires.

Une redondance externe a été utilisée pour les groupes de disques ASM, et les paramètres par défaut ont été utilisés pour les attributs des groupes de disques ASM restants. Lorsque nous avons créé les groupes de disques DATA et REDO, nous avons défini la taille de secteur sur 4 Ko afin d'optimiser les performances d'E/S sur le stockage XtremIO. Nous avons également défini la taille des blocs des journaux redo log en ligne sur 4 Ko, afin de la faire correspondre à la taille des secteurs du groupe de disques REDO.

Organisation des groupes de disques ASM pour la base de données OLTP

Plusieurs groupes de disques ASM avec redondance externe sont utilisés dans cette configuration :

- +CRS contient l'emplacement d'installation des fichiers OCR et des fichiers de vote.
- +DATA contient les fichiers de données, les fichiers de contrôle et les fichiers temporaires.
- +REDO contient les journaux redo log en ligne.

Profil de base de données et de charge applicative OLTP

Le Tableau 8 répertorie chaque profil de charge applicative de base de données OLTP de la solution.

Tableau 8. Profil de charge applicative de chaque base de données OLTP

Caractéristique du profil	Détails
Type de base de données	OLTP
Taille de la base de données	4 To
Base de données Oracle	11 gR2 RAC sur ASM
Configuration de l'instance	sga_max_size = 16 Go Remarque : puisqu'un cache de base de données plus volumineux met davantage de données en mémoire tampon, nous avons configuré un cache de mise en mémoire tampon très petit afin de générer une charge applicative d'E/S physiques stable et élevée.
Profil de charge applicative	Charge applicative OLTP simulée par SLOB
Taille de bloc de données	8 Ko

Configuration du groupe de disques ASM pour la base de données d'entrepôt décisionnel

Le Tableau 9 décrit la conception de groupe de disques ASM utilisée pour la base de données d'entrepôt décisionnel. Tout comme dans l'organisation des groupes de disques ASM OLTP, +DATA et +REDO ont été créés. Un groupe de disques ASM supplémentaire, +CSV, a été créé pour tenir lieu de partage réseau, avec le système de fichiers Oracle ACFS. Les fichiers de données plats sont stockés sur ce partage ACFS pour chargement dans la base de données.

Tableau 9. Conception du stockage pour le mappage des groupes de disques ASM dans l'environnement de base de données d'entrepôt décisionnel

Nom du dossier	Taille du volume (Go)	Nombre de volumes	Nom du groupe de disques ASM
DATA	700	8	+DATA
REDO	40	4	+REDO
CSV	200	1	+CSV
CRS	10	2	+CRS

Profil de base de données et de charge applicative DW

Le Tableau 10 décrit le profil de base de données et de charge applicative de la solution. Nous avons utilisé un générateur de charges applicatives DW pour créer une base de données d'entrepôt décisionnel et fournir ce type de charge applicative, y compris les charges applicatives de requête et de chargement des données requises dans le cadre de cette solution.

Tableau 10. Profil de base de données et de charge applicative pour chaque base de données d'entrepôt décisionnel

Caractéristique du profil	Détails
Type de base de données	Entrepôt de données
Taille de la base de données	8 To
Base de données Oracle	11gR2 RAC sur ASM
Configuration de l'instance	<ul style="list-style-type: none">• sga_target = 60 Go• pga_aggregate_target = 40 Go• db_file_multiblock_read_count = 16
Profil de charge applicative	Charge applicative DW
Source de chargement des données	Fichiers plats externes sur Oracle ACFS
Taille de bloc de données	16 Ko

Test et validation des performances

Présentation

L'environnement de test de la solution pour les systèmes OLTP et DW se composait de deux charges applicatives principales. Nous avons créé quatre machines virtuelles exécutant une base de données Oracle RAC pour la charge applicative OLTP, et quatre autres machines virtuelles destinées à l'exécution d'une base de données Oracle RAC pour la charge applicative DW dans l'environnement VMware consolidé.

Nous avons exécuté une charge applicative OLTP générée par SLOB sur la base de données afin de simuler une charge applicative OLTP réelle. Nous avons utilisé le générateur de charge applicative DW pour créer cette dernière.

Nous avons rassemblé les metrics de performances d'E/S du système (IOPS, Mo/s et latence) principalement à partir des rapports Oracle Automatic Workload Repository (AWR). En outre, nous avons collecté les metrics relatives au débit d'E/S au niveau du serveur/de la base de données et au niveau du stockage.

Remarque : en raison des chemins d'E/S et de code supplémentaires entre la couche de stockage et la couche de base de données, la latence d'E/S observée côté baie était en fait inférieure à celle observée dans les rapports AWR.

Remarques sur les résultats

Les résultats du benchmark dépendent fortement de la charge applicative, des exigences spécifiques de l'application, ainsi que de la conception et de la mise en œuvre du système. Les performances système relatives varient donc en fonction de ces facteurs et de quelques autres. En conséquence, cette charge applicative ne doit pas être utilisée comme substitut dans le cadre d'un benchmark d'application propre au client, s'il s'agit de prendre des décisions critiques concernant la planification de la capacité et/ou l'évaluation du produit.

Toutes les données de performances figurant dans ce rapport ont été obtenues dans un environnement rigoureusement contrôlé. Les résultats obtenus dans d'autres environnements d'exploitation pourraient varier sensiblement.

EMC Corporation ne reconnaît pas formellement qu'un utilisateur peut atteindre ou atteindra les mêmes performances que celles exprimées en transactions par minute.

Objectif du test

Objectifs globaux du test :

- Montrer les hautes performances atteintes en cas d'exécution des bases de données Oracle virtualisées sur des baies XtremIO.
- Montrer à quel point XtremIO simplifie les opérations de stockage dans les clusters Oracle RAC.
- Montrer le niveau soutenu des IOPS de la baie de stockage pour les charges applicatives OLTP sur une base de données Oracle RAC à quatre nœuds.
- Montrer le niveau soutenu de la bande passante d'E/S de la baie de stockage pour les charges applicatives DW/OLTP sur une base de données Oracle RAC à quatre nœuds.
- Montrer les économies d'espace considérables réalisées à l'aide de réduction des données à la volée sur
- XtremIO lorsque les volumes de production sont clonés pour les environnements de test/développement.

Scénarios de test

Les scénarios de test suivants ont été réalisés sur la solution et sont décrits dans les sections qui suivent :

Tableau 11. Scénarios de test

Scénarios de test	Nombre d'utilisateurs simultanés	Remarques
OLTP avec test 100 % requêtes	32	Charge applicative moyenne
	48	Charge applicative intensive
OLTP avec test 100 % transactions UPDATE	32	Charge applicative moyenne
	128	Charge applicative intensive
OLTP avec test 75 %/25 % QUERY/UPDATE	64	Charge applicative moyenne
	128	Charge applicative intensive
Requête DW	128	Valider le débit sur la BD RAC à 4 nœuds
Chargement des données DW	4	Valider le débit de chargement des données sur une BD RAC à 4 nœuds

Tests sur les charges applicatives pour la lecture/écriture de blocs uniques aléatoires SLOB

La solution permet de déterminer les performances système des charges applicatives OLTP Oracle sur la baie XtremIO. Nous avons créé une base de données Oracle RAC à quatre nœuds dans un environnement virtualisé VMware consolidé.

Nous avons utilisé SLOB pour générer la charge applicative, car il s'agit de la meilleure charge applicative SQL permettant de générer des E/S aléatoires physiques maximales à partir d'une plate-forme de base de données.

SLOB est un générateur de charge applicative SQL exigeante en termes d'E/S, et non un générateur d'E/S synthétiques ; il présente les caractéristiques suivantes :

- Il prend en charge le test des lectures de blocs uniques physiques et aléatoires (db file sequential read/db file parallel read).
- Il prend en charge le test des écritures de blocs uniques aléatoires (db file parallel write).
- Il prend en charge les E/S de consignation REDO extrêmes.
- Il se compose de simples requêtes PL/SQL.
- Il ne crée aucun conflit d'accès au niveau des applications.

OLTP avec test 100 % requêtes : test portant uniquement sur les requêtes de blocs uniques aléatoires SLOB

Ce test a été utilisé pour mesurer les performances lorsque plusieurs sessions « zero-think-time » SLOB simultanées exécutaient des requêtes sur la base de données RAC.

Ces sessions ont été générées par le kit d'outils SLOB : chaque utilisateur exécute des requêtes similaires pour valider les IOPS de lecture.

Méthodologie de test

Nous avons exécuté un nombre fixe de sessions simultanées, chaque utilisateur exécutant des requêtes OLTP similaires simultanément sur la base de données RAC à quatre nœuds, puis nous avons mesuré les statistiques de performances.

Procédure de test

Lors de ce test, nous avons tout d'abord exécuté une charge applicative moyenne, soit 32 sessions simultanées sur la base de données RAC à quatre nœuds, avec 8 sessions sur chaque nœud. Chaque session a exécuté la charge applicative de requêtes OLTP sur la base de données de clusters. Ensuite, nous avons exécuté une charge applicative intensive, soit 48 sessions simultanées sur la base de données RAC à quatre nœuds, avec 12 sessions sur chaque nœud. Nous avons exécuté chacune des charges applicatives trois fois afin de nous assurer qu'il ne ressortait aucune différence de taille entre elles, et surtout pour prouver que la baie XtremIO présente des performances prévisibles lorsque la même charge applicative est utilisée.

Résultats des tests

Nous avons collecté les statistiques de performances à l'aide de rapports AWR RAC. Nous nous sommes reportés à la valeur **physical read IO requests** du rapport AWR pour évaluer les statistiques relatives aux IOPS de lecture.

Le temps de réponse moyen des requêtes du processus en avant-plan a été calculé à partir des éléments **db file parallel read** et **db file sequential read**, sous **Top Timed Events** dans le rapport AWR, comme le montre la Figure 13.

Wait			Event		Wait Time			Summary Avg Wait Time (ms)				
#	Class	Event	Waits	%Timeouts	Total(s)	Avg(ms)	%DB time	Avg	Min	Max	Std Dev	Cr
*	User I/O	db file parallel read	568 666	0,00	1 450,58	2,55	49,65	2,57	2,18	2,72	0,26	
	User I/O	db file sequential read	986 577	0,00	1 060,18	1,07	36,29	1,08	0,96	1,13	0,08	
		DB CPU			583,22		19,96					
	User I/O	utl_file I/O	140	0,00	1,96	13,99	0,07	14,29	10,79	15,48	2,33	
	System I/O	control file sequential read	2 408	0,00	1,69	0,70	0,06	0,70	0,66	0,76	0,05	

Figure 13. Mesure des temps de réponse des requêtes uniquement - Rapport AWR

Remarque : lorsque le processus en avant-plan exécute les instructions SQL pour effectuer des opérations de lecture de blocs uniques sur les index, tables, segments UNDO, en-têtes de fichiers de données et autres, qui ne se trouvent pas dans le cache de mise en mémoire tampon, il émet la requête d'E/S au niveau de l'OS et attend un événement db file sequential read jusqu'à ce que le bloc requis soit renvoyé par l'OS. Pendant que les processus en avant-plan de notre test attendent les événements db file parallel read, le noyau Oracle regroupe plusieurs requêtes de lectures de blocs uniques dans un lot et émet les requêtes d'E/S au niveau de l'OS en parallèle. Le processus en avant-plan émettant les requêtes continue d'attendre jusqu'au renvoi du dernier bloc de ce lot de requêtes. Nous utilisons ces deux événements d'attente pour calculer le temps de latence moyen des opérations de lecture des processus en avant-plan.

Afin de montrer le temps de réponse moyen des E/S de la baie XtremIO du point de vue de la base de données Oracle, nous avons utilisé la formule suivante :

- Pour l'événement **db file sequential read** :
 - Le temps d'attente total est T1, soit 1 060,18 secondes.
 - Le nombre total d'événements d'attente est N1, soit 986 577.

- Pour l'événement **db file parallel read** :
 - Le temps d'attente total est T2, soit 1 450,58 secondes.
 - Le nombre total d'événements d'attente est N2, soit 568 066.
- Le temps de réponse moyen en lecture est : $(T1+T2)/(N1+N2)$, soit $(1 060,18 + 1 450,58) * 1 000 / (986 577 + 568 066)$, par conséquent, le temps de réponse moyen est de 1,615 ms.

Le Tableau 12 répertorie les IOPS et la latence d'E/S correspondante du test.

Tableau 12. IOPS de lecture et temps de réponse correspondants atteints en fonction de l'échelle du test

Metrics de performances	Données de performances					
	32 sessions			48 sessions		
	1 ^{re} passe	2 ^e passe	3 ^e passe	1 ^{re} passe	2 ^e passe	3 ^e passe
IOPS	197 288	197 601	197 128	228 092	228 011	228 599
Lecture séquentielle des fichiers bd (ms)	1,071	1,075	1,078	1,377	1,379	1,378
Lecture parallèle des fichiers bd (ms)	2,558	2,554	2,567	3,420	3,421	3,430
Latence moyenne des I/O (ms)	1,613	1,615	1,623	2,109	2,111	2,114

Remarque : le nombre d'IOPS du tableau provient de la valeur **physical write IO requests** du rapport AWR. La latence des E/S (ms) est calculée en fonction des statistiques des événements **db file sequential read** et **db file parallel read**.

Nous avons atteint un total de 197 601 IOPS de lecture et une latence moyenne de 1,615 ms avec une base de données RAC à quatre nœuds, en exécutant 32 sessions de lecture sur quatre nœuds. Lorsque nous avons augmenté la charge applicative via l'ajout de sessions de lecture, les IOPS de lecture sont passées à 228 599, avec un temps de réponse d'E/S de 2,114 ms.

OLTP avec test 100 % UPDATE : test sur les lectures/écritures de blocs uniques aléatoires SLOB

Ce test a été utilisé pour mesurer les performances lorsque plusieurs sessions « zero-think-time » SLOB simultanées exécutaient des transactions UPDATE sur la base de données RAC.

Ces sessions ont été générées par le kit d'outils SLOB : chaque utilisateur exécute des transactions UPDATE similaires pour valider les IOPS de lecture/écriture.

Méthodologie de test

Nous avons exécuté un nombre variable de sessions simultanées, chaque utilisateur exécutant des instructions SQL UPDATE SLOB similaires simultanément sur la base de données RAC à quatre nœuds, puis nous avons mesuré les statistiques de performances.

Nous avons utilisé un cache de mise en mémoire tampon (seulement 64 Mo, presque aucun bloc de données n'a été mis en cache sur les serveurs) pour chaque instance de base de données afin de pousser la charge applicative des E/S d'écritures physiques cohérentes sur le stockage back-end. La charge applicative en écriture a été générée par les instructions SQL UPDATE, qui effectuent généralement les opérations suivantes :

1. lecture des blocs de données à mettre à jour dans le cache de mise en mémoire tampon ;
2. mise à jour des lignes dans les blocs de données ;
3. validation des lignes mises à jour et déclenchement du transfert des entrées redo par LGWR vers les fichiers log en ligne.

Pendant l'exécution de la charge applicative UPDATE, le writer de base de données en arrière-plan a vidé les blocs « sales » du cache de mise en mémoire tampon et les a transférés dans des fichiers de données. Étant donné que nous avons utilisé un cache de mise en mémoire tampon très petit, les blocs de données ont été lus dans ce dernier et écrits rapidement après la mise à jour des lignes. Par conséquent, l'exécution de chaque opération UPDATE a entraîné à la fois des lectures et des écritures physiques sur la baie back-end.

Procédure de test

Lors de ce test, nous avons tout d'abord exécuté une charge applicative moyenne, soit 32 sessions simultanées sur la base de données RAC à quatre nœuds, avec 8 sessions sur chaque nœud. Chaque session a exécuté une charge applicative SQL UPDATE sur la base de données de clusters.

Ensuite, nous avons exécuté une charge applicative intensive, soit 128 sessions simultanées sur la base de données RAC à quatre nœuds, avec 32 sessions sur chaque nœud. Nous avons exécuté chacune des charges applicatives trois fois afin de nous assurer qu'il ne ressortait aucune différence de taille entre elles, et surtout pour prouver que la baie XtremIO présente des performances prévisibles lorsque la même charge applicative est utilisée.

Résultats des tests

Nous avons utilisé les statistiques de performance suivantes dans les rapports AWR pour calculer la charge applicative, comme le montre la Figure 14.

I#	Wait		Event		Wait Time			Summary Avg Wait Time (ms)			
	Class	Event	Waits	%Timeouts	Total(s)	Avg(ms)	%DB time	Avg	Min	Max	Std Dev
*	User I/O	db file sequential read	11 618 228	0,00	11 100,11	0,96	94,71	0,96	0,95	0,96	0,00
		DB CPU			978 01		8 34				
	System I/O	log file parallel write	48 686	0,00	90,51	1,86	0,77	1,86	1,84	1,88	0,02
	System I/O	db file parallel read	109 143	0,00	58,40	0,54	0,50	0,54	0,52	0,55	0,01
	Cluster	gc cr grant 2-way	37 106	0,00	10,39	0,28	0,09	0,28	0,27	0,29	0,01
	Other	reliable message	5 837	0,00	5,59	0,96	0,05	1,66	0,73	3,53	1,27

Figure 14. Mesure des temps de réponse des transactions UPDATE uniquement – Rapport AWR

- La valeur **physical write IO requests** a été utilisée pour les IOPS d'écritures physiques, qui apparaissent en tant qu'« IOPS de lecture » dans le Tableau 12.

- La valeur **physical read IO requests** a été utilisée pour les IOPS de lectures physiques entraînées par les transactions UPDATE, qui apparaissent en tant qu'« IOPS d'écriture » dans le Tableau 12. Le temps de réponse moyen en lecture a été extrait de l'événement d'attente **db file sequential read**, soit 0,96 ms dans l'exemple de rapport AWR de la Figure 14.
- La valeur **Redo size** a été utilisée pour calculer la bande passante des E/S d'écriture redo, qui apparaît en tant que « Débit redo (Mo/s) » dans le Tableau 12.
- Les statistiques associées à l'événement d'attente **log file parallel write** ont été utilisées pour la latence moyenne du processus LGWR en arrière-plan, qui apparaît en tant que « Temps de réponse LGWR (ms) » dans le Tableau 12.
- Les statistiques de l'événement d'attente **db file sequential read** ont été utilisées pour le temps de réponse moyen des lectures physiques des processus de base de données en avant-plan, qui apparaît en tant que « Temps de réponse en lecture des processus en avant-plan (ms) » dans le Tableau 12.
- Les statistiques de l'événement d'attente **db file parallel write** ont été utilisées pour le temps de réponse moyen des écritures physiques des writers de base de données en arrière-plan, qui apparaît en tant que « Temps de réponse en écriture du writer de base de données (ms) » dans le Tableau 12.

Le Tableau 13 contient les statistiques de performance du test.

Tableau 13. IOPS en lecture/écriture et temps de réponse correspondants atteints en fonction de l'échelle des sessions du test

Metrics de performances	Données de performances					
	32 sessions			128 sessions		
	1re passe	2e passe	3e passe	1re passe	2e passe	3e passe
IOPS en lecture	31 624	31 496	31 646	63 387	63 524	63 691
IOPS en écriture	33 118	32 973	33 136	66 787	66 975	67 112
IOPS agrégées (écriture + lecture)	64 742	64 469	64 782	130 174	130 499	130 803
Débit redo (Mo/s)	26	26	26	51	51	51
Temps de réponse LGWR (ms)	1,849	1,831	1,860	4,128	4,135	4,117
Temps de réponse en écriture du writer de base de données (ms)	0,535	0,531	0,535	0,643	0,642	0,637
Temps de réponse en lecture des processus en avant-plan (ms)	0,956	0,961	0,955	1,954	1,951	1,948

Comme le montre le Tableau 13, lors de l'exécution de 32 sessions simultanées sur chacun des nœuds, nous avons atteint 64 782 IOPS agrégées, y compris 33 136 IOPS en écriture et 31 646 IOPS en lecture, dans le cadre de la transaction UPDATE. En augmentant la charge applicative via l'ajout de sessions simultanées, les IOPS agrégées sont passées à 130 803, avec 63 691 IOPS en lecture et 67 112 IOPS en écriture. Le temps de réponse des E/S en lecture est passé à 1,948 ms, et celui des E/S en écriture à 0,637 ms.

En résumé, la multiplication du nombre de sessions par 4 a les conséquences suivantes :

- Les IOPS agrégées ont été doublées.
- Le débit redo (Mo/s) a été presque doublé.
- Le temps de réponse LGWR a été doublé.
- Cela a eu un impact minime sur le temps de réponse du writer de base de données.
- Le temps de réponse des processus en avant-plan a été doublé.

OLTP avec test 75 %/25 % QUERY/UPDATE test sur les blocs uniques aléatoires SLOB

Ce test a permis de mesurer les performances pendant des sessions « zero-think-time » SLOB simultanées (simulation d'utilisateurs simultanés) : 75 % d'entre elles exécutant des requêtes et 25 % d'entre elles des transactions UPDATE sur la base de données RAC. Nous avons utilisé ce test pour simuler une charge applicative réelle et valider les performances du stockage XtremIO.

Méthodologie de test

Nous avons exécuté un nombre variable de sessions simultanées : 75 % d'entre elles exécutant des requêtes OLTP similaires et les 25 % restants des transactions UPDATE similaires, simultanément, sur la base de données RAC à quatre nœuds, puis nous avons mesuré les statistiques de performance.

Procédure de test

Nous avons commencé par exécuter une charge applicative moyenne, soit 64 sessions simultanées sur la base de données RAC, avec 75 % des sessions exécutant des requêtes similaires et 25 % des instructions SQL UPDATE similaires sur la base de données de clusters. Ensuite, nous avons exécuté une charge applicative intensive, soit 128 sessions simultanées sur la base de données RAC à quatre nœuds, avec 32 sessions sur chaque nœud. Nous avons exécuté chacune des charges applicatives trois fois afin de nous assurer qu'il ne ressortait aucune différence de taille entre elles, et surtout pour prouver que la baie XtremIO présente des performances prévisibles lorsque la même charge applicative est utilisée.

Résultats des tests

Pour calculer la charge applicative, nous avons utilisé les statistiques de performance des rapports AWR, illustrés sur la Figure 15, comme nous l'avons fait lors du test 100 % transactions UPDATE.

#	Wait		Event		Wait Time			Summary Avg Wait Time (ms)				
	Class	Event	Waits	%Timeouts	Total(s)	Avg(ms)	%DB time	Avg	Min	Max	Std Dev	C...
*	User I/O	db file sequential read	7 196 550	0,00	8 153,29	1,13	70,71	1,13	1,11	1,14	0,02	
	User I/O	db file parallel read	936 985	0,00	2 416,68	2,58	20,96	2,58	2,52	2,60	0,04	
		DB CPU			1 233,61		10,70					
	System I/O	log file parallel write	244 207	0,00	254,49	1,04	2,21	1,04	1,04	1,05	0,01	
	System I/O	db file parallel read	66 278	0,00	38,73	0,58	0,34	0,58	0,56	0,60	0,02	
	Other	DFS lock handle	7 499	100,00	14,19	1,89	0,12	1,90	1,54	2,42	0,37	

Figure 15. Temps de réponse moyen - Test 75/25 QUERY/UPDATE - Rapport AWR

Le Tableau 14 répertorie les IOPS de la baie XtremIO et le temps de réponse des E/S correspondants du test. Les chiffres du Tableau 13 proviennent du rapport AWR, tout comme ceux du test 100 % UPDATE.

Tableau 14. IOPS en lecture/écriture et temps de réponse correspondants atteints en fonction de l'échelle des sessions du test

Metrics de performances	Données de performances					
	64 sessions			128 sessions		
	1re passe	2e passe	3e passe	1re passe	2e passe	3e passe
IOPS en lecture	111 797	111 886	112 124	144 911	144 977	145 205
IOPS en écriture	28 773	28 738	28 830	37 130	37 164	37 194
IOPS agrégées (écriture + lecture)	140 570	140 624	140 954	182 041	182 141	182 399
Débit redo (Mo/s)	21	21	21	28	28	28
Temps de réponse LGWR (ms)	1,045	1,037	1,042	2,119	2,101	2,121
Temps de réponse en écriture du writer de base de données (ms)	0,591	0,584	0,584	0,691	0,685	0,690
Lecture séquentielle des fichiers bd (ms)	1,138	1,136	1,133	1,747	1,748	1,746
Lecture parallèle des fichiers bd (ms)	2,585	2,583	2,579	4,298	4,302	4,300
Temps de réponse moyen en lecture des processus en avant-plan (ms)	1,304	1,303	1,299	2,039	2,040	2,039

Remarque : pour connaître la source des metrics de performances de ce tableau, reportez-vous aux sections précédentes.

En ce qui concerne les charges applicatives mixtes incluant les requêtes et les transactions UPDATE, les résultats de nos tests montrent que le stockage XtremIO peut maintenir un haut niveau d'IOPS. Dans notre test 75/25 QUERY/UPDATE de 64 sessions simultanées, les IOPS en lecture ont atteint 112 124 et les IOPS en écriture 28 830, maintenant la latence des E/S en lecture à près de 1 ms et la latence des E/S en écriture à près de 0,59 ms. En faisant passer la charge applicative à 128 sessions simultanées, les IOPS agrégées sont passées à 182 399, avec 145 205 IOPS en lecture et 37 194 IOPS en écriture. Le temps de réponse en lecture est passé à 2,039 ms, et celui du writer de base de données à 0,69 ms.

Test sur la charge applicative des requêtes d'entrepôt décisionnel

Le générateur de charge applicative d'entrepôt décisionnel nous a fourni une charge applicative de test d'entrepôt décisionnel Oracle qui nous a permis de tester et de valider les performances de charges applicatives DW Oracle types sur la plate-forme de stockage XtremIO.

Le schéma du kit présentait 12 tables, y compris deux tables des faits : sales et returns. Les tables restantes tenaient lieu de tables des dimensions. Les tables des faits ont été partitionnées par date et sous-partitionnées en fonction du hachage sur leur clé de jointure. La base de données faisait 3,8 To. Plusieurs utilisateurs simultanés ont exécuté une série de requêtes ordinaires sur les données. Le débit a été mesuré pendant le test.

Méthodologie de test

Ce test a permis de mesurer les performances pendant l'exécution d'un nombre fixe d'utilisateurs simultanés, chaque utilisateur exécutant des requêtes similaires. Ces utilisateurs simultanés ont été générés par le générateur de charge applicative DW. La valeur « direct path read » a été choisie pendant l'exécution, pour toutes les analyses en série des tables complètes de ce test.

Procédure de test

Ce test a impliqué l'exécution de 32 utilisateurs simultanés sur chaque nœud, soit un total de 128 utilisateurs simultanés sur la base de données RAC à quatre nœuds, chaque session exécutant une charge applicative de requêtes DW similaire, avec les scripts du générateur de charge applicative DW.

Résultats des tests

Pour évaluer le débit des requêtes (Go/s), nous avons utilisé l'élément **physical read bytes** du rapport AWR, comme le montre la Figure 16.

physical read IO requests	9,326,610	41,065.10	3,108,870.00
physical read bytes	1,218,506,964,992	5,365,091,054.29	406,168,988,330.67
physical read total IO requests	9,328,229	41,072.23	3,109,409.67

Figure 16. Débit des requêtes – Test de la charge applicative de requêtes DW – Rapport AWR

Le rapport AWR montre que pendant les tests, le débit a atteint 5 Go/s lorsque nous avons exécuté la charge applicative DW sur la base de données RAC à quatre nœuds.

Test de chargement des données d'entrepôt décisionnel

Les entrepôts décisionnels des entreprises modernes requièrent des chargements de données fréquents et volumineux tout au long de la journée. Le besoin d'un entrepôt décisionnel fonctionnant en permanence ne laisse plus beaucoup de temps aux administrateurs BD pour le chargement des données. Par conséquent, il est important de simuler l'impact des processus d'extraction, de transformation et de chargement des données (ETL) sur les performances de la base de données.

Ce scénario de test montre l'incidence des processus ETL sur la base de production et enregistre les données de performances, plus particulièrement le débit (Mo/s d'écritures physiques totales), pendant la charge ETL.

Nous avons eu recours à des tables externes Oracle utilisant le pilote d'accès ORACLE_LOADER pour charger les données de ces tables vers des tables internes. Les données provenaient de fichiers de données plats.

Ce scénario de test montre l'évolutivité du débit lors du chargement des données à partir de tables externes situées dans le système de fichiers Oracle ACFS vers la base de données.

Méthodologie de test

Ce test montre le débit observé dans la baie XtremIO avec la configuration du stockage sur disque utilisée dans cette solution. Une session est exécutée sur chaque serveur (pour un total de quatre) pour le chargement des données sur la base de données RAC à quatre nœuds. Chaque session a exécuté une charge applicative ETL similaire en chargeant des fichiers CSV plats dans la base de données.

Procédure de test

Ce test a démontré les performances des modules X-Brick en chargeant des données à partir de tables externes. Il a impliqué l'exécution d'un utilisateur sur chaque nœud de la base de données RAC qui en compte quatre, pour charger les données à partir d'une table externe, équivalant à quatre sessions chargeant les données simultanément.

Chaque session a chargé un fichier CSV de 120 Go. Ce fichier CSV se situait dans le système de fichiers Oracle ACFS. La table externe a été créée comme suit :

```
create table sales_ext (
  id integer,
  ...)
organization external(
  type oracle_loader
  default directory EXT_DIR
  access parameters (fields terminated by "|")
  location ('sales.csv'))
parallel reject limit unlimited;
```

Les données ont été chargées à partir de la table externe comme suit :

```
alter session enable parallel dml;
alter table sales parallel;
alter table sales_ext parallel;
insert into /**+ append */ sales select * from sales_ext;
```

Remarque : la table **sales** présente la même structure que la table **sales_ext**. Les données sont chargées directement à l'aide de l'indicateur **append**, et plusieurs esclaves parallèles sont utilisés pour le chargement des données.

Résultats des tests

Pour évaluer le débit (To/heure), nous avons utilisé l'élément **physical write bytes** du rapport AWR, comme le montre la Figure 18.

physical write IO requests	282,266	4,205.96	94,088.67
physical write bytes	73,635,045,376	1,097,212,611.86	24,545,015,125.33
physical write total IO requests	289,038	4,306.86	96,346.00

Figure 17. Débit de chargement des données – Test de chargement des données DW - Rapport AWR

Le rapport AWR montre que pendant le test de chargement des données, le débit a atteint près de 1 023 Go/s, soit 3,59 To/heure, lorsque nous avons exécuté la charge applicative de chargement des données sur la base de données RAC à quatre nœuds.

Conclusion

Résumé

La solution Oracle Database 11g avec EMC XtremIO et VMware vSphere permet une véritable évolutivité des instances Oracle consolidées et virtualisées sur plusieurs charges applicatives. Cette solution peut servir de base à un système pouvant évoluer de manière flexible, prévisible et pratiquement linéaire.

Elle fournit une infrastructure haute évolutivité pour les applications et leur permet de gérer une charge applicative toujours plus importante, avec des performances prévisibles. Côté serveur, la solution est évolutive et peut s'adapter à l'ajout de ressources de traitement supplémentaires, telles que des CPU, de la mémoire, des ports d'adaptateur HBA et des ports front-end. Côté stockage, l'architecture scale-out de type actif/actif de la baie XtremIO permet de faire évoluer la capacité et les débits de manière linéaire (IOPS et Mo/s) tout en maintenant une latence supérieure et cohérente. Ces facteurs indiquent que cette solution est parfaitement adaptée aux environnements de bases de données OLTP et DW.

Avec cette solution, vous pouvez également profiter d'une utilisation des ressources efficace grâce à la virtualisation, tout en permettant la consolidation des bases de données et des applications. Vous pouvez faire évoluer la capacité et les fonctionnalités de traitement de façon indépendante, sans que vos efforts ne soient entravés par les limites imposées par une appliance dédiée. En cas de modification, quelle qu'elle soit, au niveau des applications, des logiciels de base de données ou des autres logiciels, la pile de la solution peut s'adapter aux demandes changeantes imposées par vos besoins métiers. En retour, de nouvelles approches des charges applicatives peuvent être adoptées, comme l'analytique en temps réel, grâce à la consolidation des instances de production et de reporting.

Conclusions

La solution présente les avantages suivants :

- configuration d'XtremIO rapide et simple sans réglage du stockage requis ;
- entrepôt décisionnel à large bande passante avec plusieurs utilisateurs simultanés et charge applicative aléatoire de blocs de taille volumineuse ;
- performances et flexibilité améliorées avec une latence réduite et des IOPS plus élevées ;
- bases de données en cluster virtualisées avec E/S élevées, comme Oracle RAC ;
- économies importantes en termes d'encombrement du stockage via l'utilisation de la déduplication à la volée XtremIO ;
- protection/tolérance aux pannes élevée pour les systèmes en cluster haute disponibilité tels qu'Oracle RAC et vSphere HA.

Références

Livres blancs

Pour obtenir des informations complémentaires, consultez les documents EMC suivants. L'accès à ces documents dépend de vos informations d'identification. Si vous ne parvenez pas à accéder à un document, contactez un responsable de compte EMC.

- *EMC Infrastructure for High Performance Microsoft and Oracle Database Systems*
- *Introduction to the EMC XtremIO Storage Array*

Documentation VMware

Pour obtenir des informations complémentaires, consultez les documents VMware suivants.

- Windows XP Deployment Guide
- Oracle Databases on VMware Best Practices Guide
- Base de connaissances VMware : Large-scale workloads with intensive I/O patterns might require queue depths significantly greater than Paravirtual SCSI default values (2053145)

Documentation Oracle

Pour obtenir des informations complémentaires, consultez le document *Oracle Database Client Installation Guide 11g Release 2 (11.2) for Linux*.

Autre documentation

Les liens suivants contiennent des informations complémentaires sur les snapshots XtremIO.

- Introduction to XtremIO Snapshots White Paper
<http://www.xtremio.com/introduction-to-xtremio-snapshots-white-paper>
- XtremIO Overview: Snapshot Technologies for Oracle Database
<https://community.emc.com/docs/DOC-30738>