

Dell PowerMax : réduction des données

Compression et déduplication à la volée

Juillet 2022

H19254

Livre blanc

Résumé

Les plates-formes de stockage PowerMax intègrent plusieurs techniques de réduction des données, notamment la compression et la déduplication à la volée. Elles comprennent également des fonctionnalités de détection des schémas et de positionnement efficace des données afin d'offrir un excellent équilibre entre performances et efficacité.

Dell Technologies

Copyright

Les informations contenues dans cette publication sont fournies « en l'état ». Dell Inc. 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 distribution de tout logiciel décrit dans cette publication nécessitent une licence logicielle en cours de validité.

Copyright © 2022 Dell Inc. ou ses filiales. Tous droits réservés. Dell Technologies, Dell, EMC, Dell EMC et les autres marques citées sont des marques de Dell Inc. ou de ses filiales. Intel, le logo Intel, le logo Intel Inside et Xeon sont des marques commerciales d'Intel Corporation aux États-Unis et/ou dans d'autres pays. Les autres marques peuvent être la propriété de leurs détenteurs respectifs. Publié en France en juillet 2022 H19254.

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

Sommaire

- Synthèse..... 4
- Réduction des données..... 5
- Utilisation des ressources système 9
- Gestion et surveillance..... 9
- Services de données pris en charge..... 17
- Conclusion 18
- Références 19

Synthèse

Présentation

La réduction des données avec le système Dell PowerMax optimise l'efficacité du système en combinant la compression à la volée, la déduplication à la volée et la détection des schémas. L'utilisation de ces techniques de réduction des données permet aux utilisateurs d'économiser énormément de capacité. La réduction des données compresse les données et élimine les copies redondantes. Ce livre blanc décrit le fonctionnement de la réduction des données dans les systèmes PowerMax et explique comment créer des rapports à l'aide d'applications de gestion Dell telles que Unisphere for PowerMax, Solutions Enabler et le logiciel Mainframe Enabler.

Révisions

Date	Description
Juillet 2022	Version originale

Vos commentaires sont les bienvenus.

Dell Technologies et les auteurs vous invitent à leur faire part de vos commentaires concernant ce document. Contactez l'équipe Dell Technologies par [e-mail](#).

Auteur : Robert Tasker

Remarque : Trouvez des liens vers d'autres documents traitant du même sujet dans l'[InfoHub sur PowerMax et VMAX](#).

Réduction des données

Présentation

La réduction des données associe la compression à la volée, la déduplication à la volée, la détection des schémas, le positionnement efficace des données et l'apprentissage automatique (ML). Cette combinaison permet d'obtenir un système sur lequel les utilisateurs peuvent écrire une quantité de données hôte supérieure à la quantité totale de capacité physique disponible, tout en atteignant les performances qu'ils peuvent attendre d'un système de stockage d'entreprise. Cette fonction est activée par défaut et vous pouvez l'activer ou la désactiver au niveau du groupe de stockage. En outre, tous les services de données disponibles dans les systèmes PowerMax 2500 et 8500 sont pris en charge. Cette prise en charge s'applique également à l'émulation CKD, bien qu'elle ne couvre pas la déduplication pour CKD.

La compression réduit la taille des données, tandis que la déduplication stocke les données sous la forme d'une instance unique. La détection de schémas inclut une fonction d'allocation non nulle qui empêche de stocker les chaînes de zéros consécutifs avec les données compressées. La compression, la déduplication et la détection des schémas sont exécutées à l'aide d'une assistance matérielle intégrée au système afin de réduire le temps système nécessaire à l'exécution de ces fonctions. L'apprentissage automatique identifie les données les plus actives stockées sur le disque et empêche leur réduction afin de garantir des performances optimales. Le positionnement efficace des données utilise une fonction appelée **compactage** qui stocke stratégiquement les données afin de minimiser les pertes d'espace et évite d'avoir à exécuter les fonctions de nettoyage de la mémoire ou de défragmentation.

Réduction basée sur l'activité

La réduction basée sur l'activité (ABR, Activity Based Reduction) réduit les coûts de performances liés à la décompression des données auxquelles les utilisateurs accèdent fréquemment. Cette fonction permet de stocker sur le système jusqu'à 20 % des données les plus actives sans qu'elles ne soient compressées. Cette capacité est avantageuse pour le système, car elle réduit la latence de performances résultant d'une décompression constante des données fréquemment consultées. Pour déterminer quelles données sont les plus actives, le système utilise des algorithmes ML qui traitent les statistiques d'E/S. Cette tâche permet de maintenir un environnement équilibré et optimal pour économiser et améliorer les performances de la réduction des données.

Compression

La compression réduit au minimum les charges applicatives d'écriture entrantes pour consommer le moins de capacité possible. Les données sont compressées au moment de leur transmission au matériel de réduction des données qui utilise l'algorithme de compression GZIP. Une fois transmises au matériel de réduction des données, les données sont divisées en quatre sections, lesquelles sont compressées en parallèle pour garantir une efficacité optimale. La somme des quatre sections correspond à la taille réduite finale des données stockées sur le disque. Cette capacité fournit un accès granulaire aux données ainsi réduites. Seules les sections qui contiennent les données demandées pour les demandes de lecture ou d'écriture partielles sont traitées, étant donné que chaque section peut être gérée indépendamment.

Déduplication

La déduplication est une méthode d'économie de capacité qui identifie des copies identiques de données et stocke une instance unique de chaque copie. Certains composants de la déduplication sont nécessaires pour permettre de réaliser de réelles économies de capacité.

- **ID de hachage** : l'ID de hachage est un identifiant unique pour les données entrantes qui est utilisé pour déterminer si une relation de déduplication est nécessaire. Le système utilise un algorithme SHA-256 pour générer l'ID de hachage.
- **Table des ID de hachage** : les tables de hachage représentent une allocation de mémoire système distribuée entre les directeurs système. Ces tables répertorient les ID de hachage utilisés par le processus de déduplication. Les entrées de la table sont utilisées pour déterminer s'il existe une relation de déduplication, ou si une nouvelle entrée est nécessaire et si les données peuvent être stockées sur le disque.
- **Objet de gestion de déduplication** : le DMO (Dedupe Management Object) est un objet de 64 octets au sein de la mémoire système qui n'existe qu'en cas de relation de déduplication. Ces objets stockent et gèrent les pointeurs entre les appareils

front-end et les données dédupliquées qui consomment de la capacité back-end dans la baie. Les DMO gèrent les pointeurs pour les données dédupliquées entre les appareils front-end et les données stockées sur le disque. Cela permet également de gérer la table de hachage dans laquelle les identifiants de hachage sont stockés lorsque des relations de déduplication existent.

La déduplication est effectuée à l'aide du même matériel de réduction des données que la compression, et un ID de hachage unique est généré lorsque les données sont traitées par le matériel. L'ID de hachage est ensuite comparé à la table des ID de hachage afin de rechercher le même ID. Lorsqu'une correspondance est trouvée, les données ne sont pas stockées sur le disque et un partage de déduplication est créé. Les pointeurs sont définis entre le volume front-end et l'ID unique dans la table des ID de hachage. Les pointeurs relient l'instance unique des données stockées sur le disque au volume, ce qui permet par la suite d'accéder aux données. Le DMO gère les pointeurs entre les données, les volumes front-end qui accèdent aux données et la table des ID de hachage. Lorsqu'il n'y a aucune correspondance dans la table des ID de hachage, une nouvelle entrée y est ajoutée pour les futures comparaisons d'ID de hachage.

Algorithme de déduplication

Les systèmes PowerMax utilisent l'algorithme de hachage SHA-256 implémenté dans le matériel de réduction des données pour trouver des données dupliquées. Les données sont ensuite stockées sous la forme d'une instance unique partagée entre plusieurs sources. Ce processus améliore l'efficacité des données tout en préservant leur intégrité.

L'algorithme SHA-256 génère un code de 32 octets pour chaque bloc de données de 32 Ko. Exemple : un système avec 1 PB de données écrites avec 5 % de mise à jour par jour. Sur 1 million d'années de fonctionnement, il y a une probabilité de 20 % d'une collision de hachage. Chaque piste de 128 Ko étant gérée sous la forme de quatre blocs de 32 Ko, une collision de hachage est nécessaire sur les quatre blocs de la même piste de 128 Ko pour produire une collision de hachage réelle. Les chances d'avoir quatre collisions de blocs simultanées sont infinitésimales (moins de 1 % de chance pendant un billion d'années de fonctionnement). En outre, lorsqu'une correspondance est trouvée au cours de la phase de comparaison de la déduplication, une comparaison octet par octet est effectuée. Cette comparaison permet de confirmer qu'il existe une correspondance avant de mettre à jour les tables et de définir les pointeurs pour autoriser l'accès aux données.

Compactage

Le positionnement des données est effectué à l'aide d'un processus appelé « compactage ». La compression place intuitivement les données réduites ou non réduites au meilleur emplacement possible sur le disque. L'opération de stockage des données sur le disque utilise des objets d'écriture. Chaque objet comprend 6 Mo de capacité contiguë de périphérique de données back-end sur l'ensemble des disques configurés dans le système. Les objets d'écriture sont alignés sur des limites de 1 Ko et sont consommés de manière séquentielle en une seule utilisation. Les objets d'écriture sont répartis sur des bandes entières pour tous les types RAID pris en charge afin d'optimiser les écritures. Chaque objet prend en charge les données réduites ou non pour les émulations FBA et CKD.

- **Objet d'écriture FBA** : un objet d'écriture non réduit se compose de 48 pistes FBA. Un objet d'écriture réduit se compose de 1 000 pistes réduites. Les entrées réduites pour les objets d'écriture vont de 1 Ko à 96 Ko.
- **Objet d'écriture CKD** : un objet d'écriture non réduit se compose de 108 pistes CKD. Un objet d'écriture réduit se compose de 1 000 pistes réduites. Les entrées réduites pour les objets d'écriture vont de 1 Ko à 52 Ko.

Extended Data Compression

Les systèmes PowerMax 2500 et 8500 incluent une fonction supplémentaire appelée « Extended Data Compression » (EDC), qui compresse les données déjà compressées afin d'économiser davantage de capacité. Cette tâche consiste à identifier les données qui n'ont pas été consultées pendant une durée prolongée. Les facteurs qui font des données des candidates idéales pour l'EDC sont répertoriés ci-dessous :

- Les données appartiennent à un groupe de stockage activé pour la réduction des données.
- Les données n'ont pas été consultées pendant 30 jours.
- Les données ne sont pas déjà compressées par EDC.

Les données éligibles à l'EDC sont compressées à l'aide de l'algorithme Def9_128_SW afin de réduire davantage la capacité utilisée pour stocker les données. Ce processus automatisé est exécuté en arrière-plan dans le système. Il est encore possible d'économiser sur la capacité à l'aide du taux de compression atteint au niveau du groupe de stockage. EDC est uniquement disponible avec les baies de stockage PowerMax.

CKD compression

La réduction basée sur l'activité (ABR, Activity Based Reduction) réduit les coûts de performances liés à la décompression des données auxquelles les utilisateurs accèdent fréquemment. Cette fonction permet de stocker sur le système jusqu'à 20 % des données les plus actives sans qu'elles ne soient compressées. Ce résultat profite au système, car il élimine l'impact négatif sur les performances résultant d'une décompression constante des données fréquemment consultées. Pour déterminer le niveau d'activité des données, le système utilise des algorithmes ML qui traitent les statistiques collectées entre les E/S entrantes et les appareils front-end. Cette opération permet au système de maintenir un équilibre entre les ressources du système, ce qui offre un environnement optimal pour les économies et les performances de réduction des données.

La compression réduit au minimum les charges applicatives d'écriture entrantes pour consommer le moins de capacité possible. Les données sont compressées au moment de leur transmission au matériel de réduction des données intégré au système, lequel utilise l'algorithme de compression GZIP. Une fois transmises au matériel de réduction des données, les données sont divisées en quatre sections, lesquelles sont compressées en parallèle pour garantir une efficacité optimale du matériel. La somme des quatre sections correspond à la taille réduite finale des données stockées sur le disque. Ce résultat fournit un accès granulaire aux données réduites en cas de demande de lecture ou d'écriture partielle. Seules les sections qui contiennent les données demandées sont traitées, étant donné que chaque section peut être gérée indépendamment.

Le positionnement des données est effectué à l'aide d'un processus appelé « compactage ». La compression place intuitivement les données réduites ou non réduites au meilleur emplacement possible sur le disque. L'opération de stockage des données sur le disque utilise des objets d'écriture. Chaque objet comprend 6 Mo de capacité contiguë de périphérique de données back-end sur l'ensemble des disques configurés dans le système. Les objets d'écriture sont alignés sur des limites de 1 Ko et sont consommés de manière séquentielle en une seule utilisation. Les objets d'écriture sont répartis sur des bandes entières pour tous les types RAID pris en charge afin d'optimiser les écritures. Chaque objet prend en charge les données réduites ou non réduites. Un objet d'écriture non réduit se compose de 108 pistes CKD. Un objet d'écriture réduit se compose de 1 000 pistes réduites. Les entrées réduites pour les objets d'écriture vont de 1 Ko à 52 Ko.

Flux d'E/S de réduction des données

Toutes les E/S sont transmises via le cache, puis traitées par le système. Les actions de réduction des données sont effectuées après réception des données par le système, avant qu'elles ne soient placées sur le disque. L'utilisation d'un processus à la volée nécessite des vérifications supplémentaires au sein du débit d'E/S où la réduction des données s'applique. Le système utilise ces vérifications pour déterminer si les données entrantes doivent passer par le matériel de réduction des données ou non. Les données entrantes pour un groupe de stockage, avec réduction des données activée, suivront le flux de réduction des données. Cependant, en raison de la

fonction de réduction basée sur l'activité (ABR), les données actives d'un groupe de stockage avec la réduction des données activées vont ignorer le débit de réduction des données pour l'optimisation des performances. Les données non compressées dues à ABR peuvent être compressées ultérieurement et déplacées vers un pool de compression. Les données d'un groupe de stockage dont la réduction des données est désactivée ignorent le débit de réduction des données et sont écrites sur le système sans compression.

Il existe quelques types d'E/S différents à prendre en compte : la lecture, l'écriture et la mise à jour en écriture.

- **Lecture** : demande d'accès aux données qui remplissent déjà la baie.
- **Écriture** : E/S entrantes qui consommeront de l'espace disque.
- **Mise à jour en écriture** : E/S entrantes qui peuvent modifier les données allouées à l'espace disque sur la baie.

La figure suivante décrit le chemin suivi par les E/S, qui est déterminé par les caractéristiques du jeu de données ou du groupe de stockage associé.

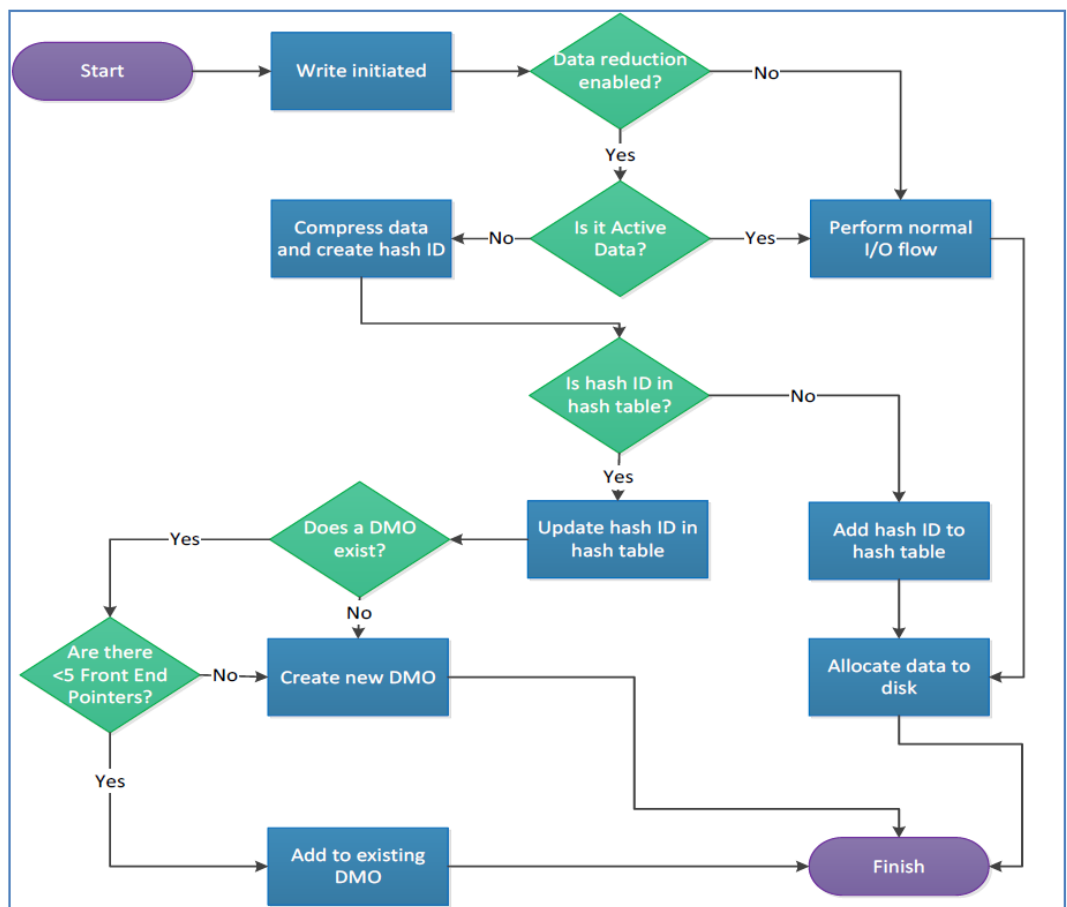


Figure 1. Débit d'E/S de réduction des données pour les systèmes de stockage d'entreprise PowerMax

Utilisation des ressources système

Présentation

La capacité et la mémoire sont les deux principales ressources configurées dans chaque système PowerMax. La capacité est classée en plusieurs catégories : physique, effective et provisionnée. La mémoire est classée en ressources système.

Les ressources de mémoire prennent en charge les structures de métadonnées pour la capacité provisionnée ainsi que pour la capacité physique. La capacité effective disponible est liée à la quantité de capacité physique, à la quantité de ressources système disponibles et à la réductibilité des données écrites sur le système. Les données écrites qui sont hautement réductibles consomment moins de capacité physique, ce qui se traduit par davantage de capacité effective. L'inverse est également vrai : les données écrites qui ne sont pas réductibles peuvent entraîner moins de capacité effective disponible. Les informations décrites dans la section [Réduction des données](#) (capacité, ressources système) sont disponibles dans les applications de gestion utilisées pour les systèmes PowerMax 2500 et 8500, à savoir Unisphere for PowerMax, Solutions Enabler et le logiciel Mainframe Enabler. Unisphere for PowerMax est une interface utilisateur qui fournit des données sous forme de graphiques, de tableaux et de listes. Solutions Enabler est une interface de ligne de commande standard qui fournit les mêmes données, mais pas sous la forme de graphiques et de tableaux. Mainframe Enablers est une suite de composants qui surveillent et gèrent les systèmes Dell Storage dans un environnement mainframe. Les images affichées dans les sections suivantes de ce livre blanc illustrent l'application Unisphere for PowerMax qui gère un système PowerMax 2500 ou 8500.

Capacité physique

La capacité physique correspond à la quantité d'espace disque configurée dans le système en fonction des disques installés et de la protection RAID appliquée. Dans une configuration où la réduction des données n'est pas utilisée, la capacité physique correspond à la quantité totale de capacité disponible pour les données hôte. Par exemple, un système affichant 100 To de capacité physique indique qu'il peut accueillir 100 To de données hôte qui n'utilisent pas la réduction des données.

Capacité effective

La capacité effective est la quantité d'espace disponible lorsque la réduction des données est utilisée. La quantité totale lors de l'installation initiale dépend de la quantité de mémoire configurée dans le système ; elle est fondée sur des taux de réduction des données par défaut de 4:1 (3:1 pour l'émulation CKD).

Par exemple, ce même système avec 100 To de capacité physique affichera 400 To de capacité effective. Cette valeur de 400 To constitue un point de départ pour la capacité effective ; elle change à mesure que les données sont écrites sur le système et que la réduction des données est appliquée.

Capacité provisionnée

La capacité provisionnée est la représentation de la capacité disponible sous la forme d'appareils créés et présentés aux hôtes et aux applications qui ont l'intention de consommer de la capacité physique ou effective dans le système.

Gestion et surveillance

Présentation

Unisphere for PowerMax est une interface utilisateur qui permet de gérer et surveiller la capacité et l'utilisation des ressources du système. Au niveau du système, des informations sur l'utilisation de la capacité, la réduction des données et les ressources système s'affichent dans le tableau de bord Capacity. À partir du tableau de bord Capacity, les utilisateurs peuvent accéder à des écrans qui fournissent des informations sur la capacité effective et la capacité provisionnée, la capacité des snapshots, la réduction des données et les ressources système.

Tableau de bord Capacity

Dans Unisphere for PowerMax, plusieurs écrans fournissent des informations sur l'utilisation de la capacité.

Le tableau de bord principal affiche un graphique interactif qui indique l'utilisation effective de la capacité et la réduction des données au fil du temps. Cet affichage indique l'historique d'utilisation de la capacité effective et la relation entre le taux de réduction des données et la capacité effective. Ces informations peuvent être utilisées pour surveiller et suivre les tendances d'utilisation de la capacité effective par rapport au taux de réduction des données affiché. Les systèmes PowerMax 2500 ou 8500 peuvent être configurés avec l'émulation FBA et CKD au sein du même pool de ressources de stockage, mais le graphique d'historique est spécifique à la vue d'émulation sélectionnée.

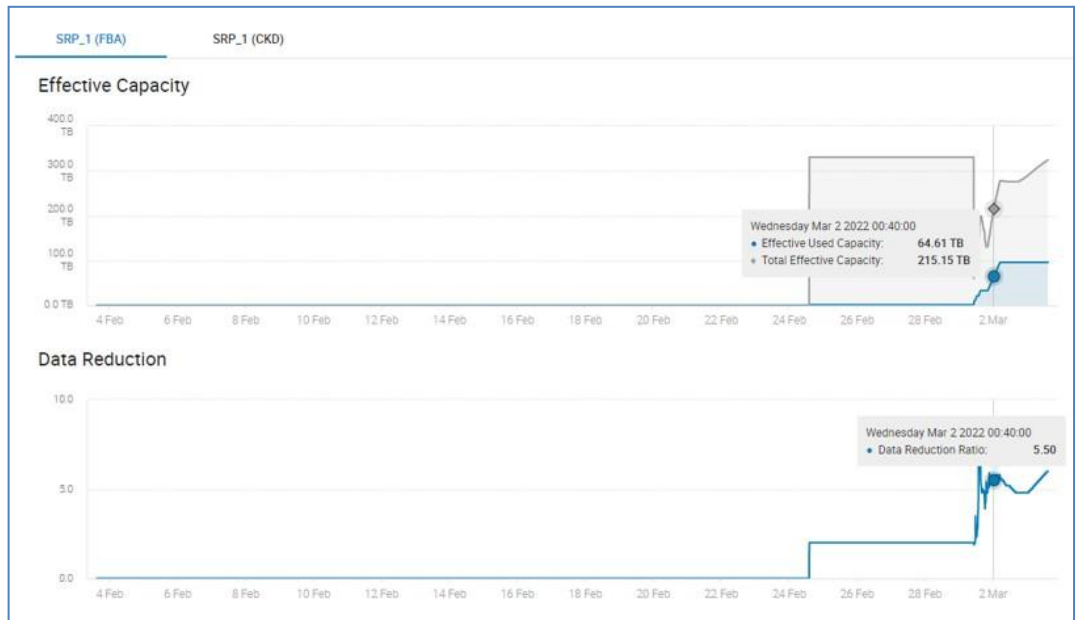


Figure 2. Graphique d'historique du tableau de bord Capacity indiquant la capacité effective et la réduction des données pour l'émulation FBA

Le tableau de bord principal fournit également, sous la forme de graphiques à barres, des données concernant la capacité provisionnée, la capacité effective, l'utilisation des snapshots et la réduction des données. Chaque section peut être étendue à un affichage plus détaillé qui présente des données plus granulaires pour chaque élément.

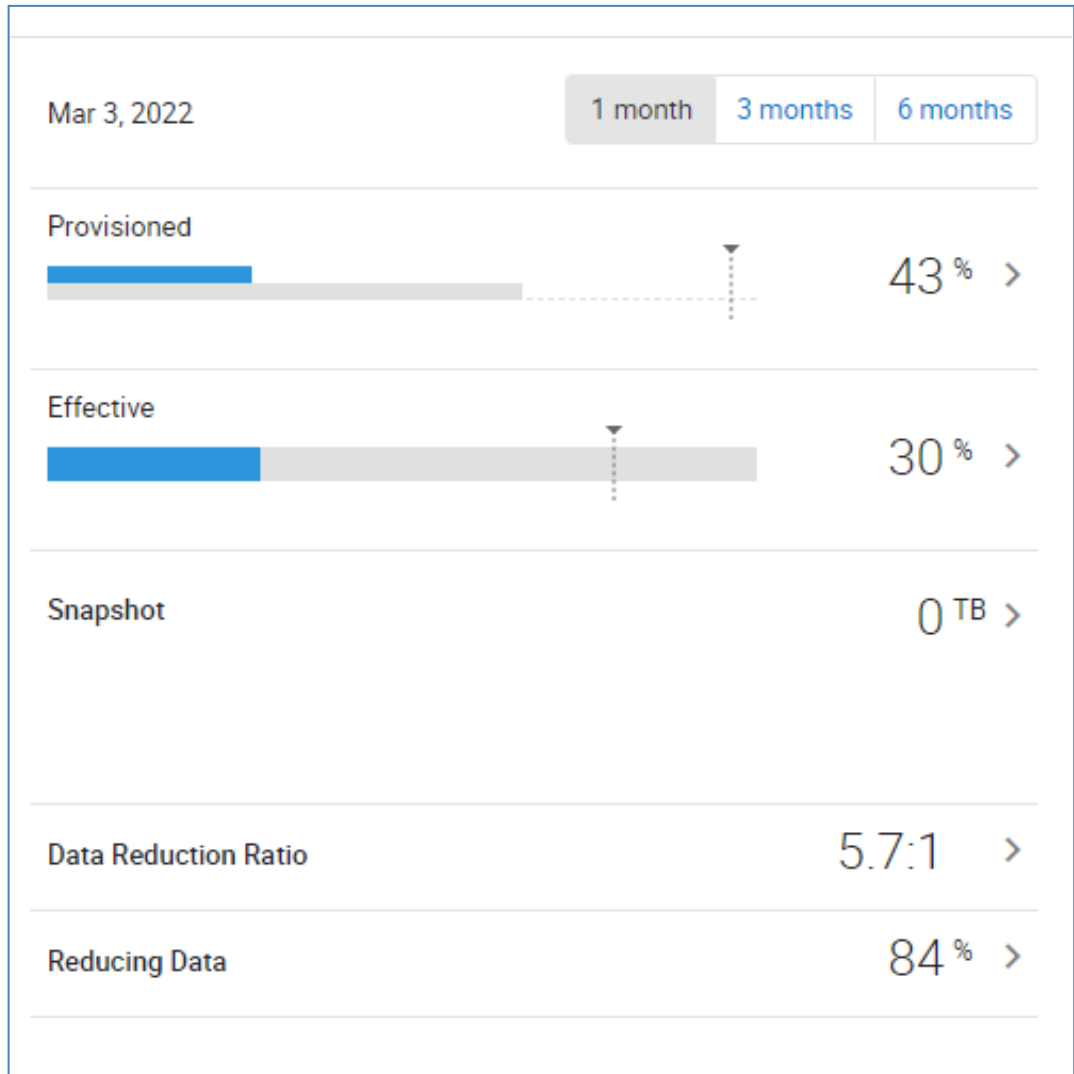


Figure 3. Graphiques à barres du tableau de bord Capacity indiquant la capacité provisionnée, la capacité effective, l'utilisation des snapshots et la réduction des données

dynamique

La capacité provisionnée correspond à la quantité de capacité provisionnée sous la forme d'appareils présentés aux hôtes et aux applications en tant que capacité disponible. Le suivi de la capacité provisionnée est effectué à l'aide de deux métriques : SRP Capacity et System Resources.

- SRP Capacity** indique la quantité de capacité provisionnée sous la forme d'une valeur de To provisionnés et la quantité disponible de capacité effective. La quantité de capacité SRP effective est initialement basée sur des taux de réduction de données par défaut de 4:1 (3:1 pour l'émulation CKD). Cette quantité est calculée à partir de la quantité de capacité physique configurée dans le système. À mesure que les utilisateurs créent des appareils, la capacité provisionnée augmente. La valeur en pourcentage indiquée correspond au pourcentage d'abonnement et est calculée à l'aide des quantités de capacité provisionnée et de capacité effective.

- La métrique **System Resources** représente la quantité de mémoire disponible pour prendre en charge la capacité provisionnée sous la forme de métadonnées. La valeur totale présentée ne change que si du cache est ajouté au système. La quantité utilisée représente précisément ce qui est utilisé. La quantité libre indique la quantité de capacité provisionnée supplémentaire que le système peut prendre en charge. Au fur et à mesure que les utilisateurs créent des appareils, la quantité utilisée augmente. Lorsqu'il existe une variation entre les deux valeurs, les ressources de mémoire sont utilisées pour prendre en charge la réduction des données ou d'autres fonctionnalités qui utilisent la mémoire, telles que la création de snapshots d'appareils existants.

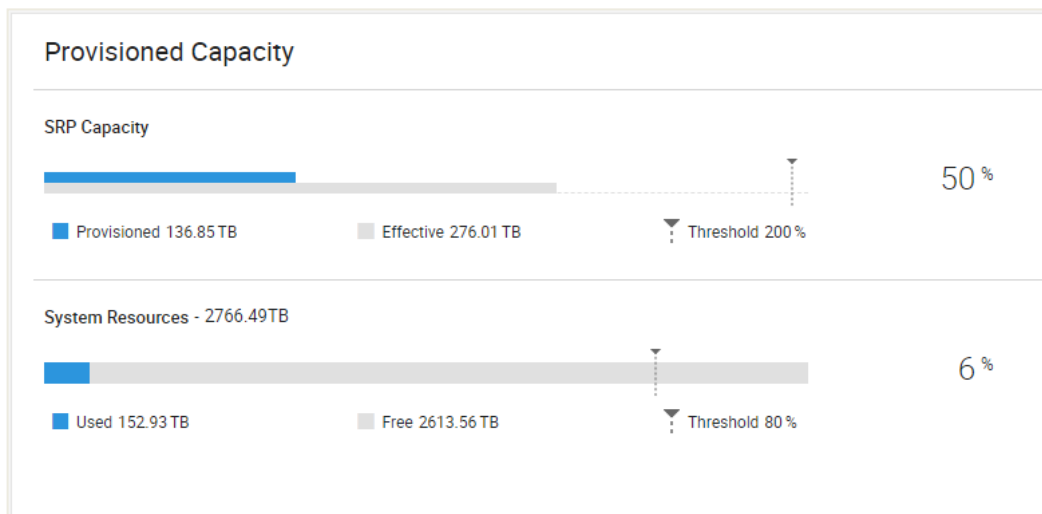


Figure 4. Utilisation des ressources système dans l'affichage de la capacité provisionnée

Réelle

La capacité effective représente la quantité de capacité disponible pour l'utilisateur en fonction des attentes d'économies liées à l'utilisation de la réduction des données. L'affichage de la capacité effective fournit une vue détaillée des ressources physiques et effectives disponibles. Cette vue est divisée en trois sections : Physical Capacity, Effective Capacity Resources et Effective Capacity Usage.

- La section **Physical Capacity** indique la quantité de capacité physique disponible sur les disques durs configurés dans le système. Les quantités indiquées correspondent aux valeurs obtenues après l'application du formatage et de la protection RAID. La valeur affichée est la quantité de capacité que le système peut prendre en charge pour les données hôte lorsque la réduction des données n'est pas utilisée.
- La section **Effective Capacity Resources** indique les valeurs réalisables en fonction de l'utilisation actuelle des ressources système. La valeur des ressources de capacité effective indiquée s'ajuste par rapport aux économies actuelles liées à la réduction des données et compte tenu de l'utilisation de la capacité physique et de la capacité effective.
- La section **Effective Capacity Usage** affiche la quantité actuelle de capacité effective disponible en fonction de l'utilisation des ressources système et des économies actuelles liées à la réduction des données. La valeur affichée dans le graphique circulaire correspond à la valeur actuelle de capacité effective disponible. Les valeurs présentées à droite divisent l'utilisation en trois catégories : Snapshot Used, User Used et Free.

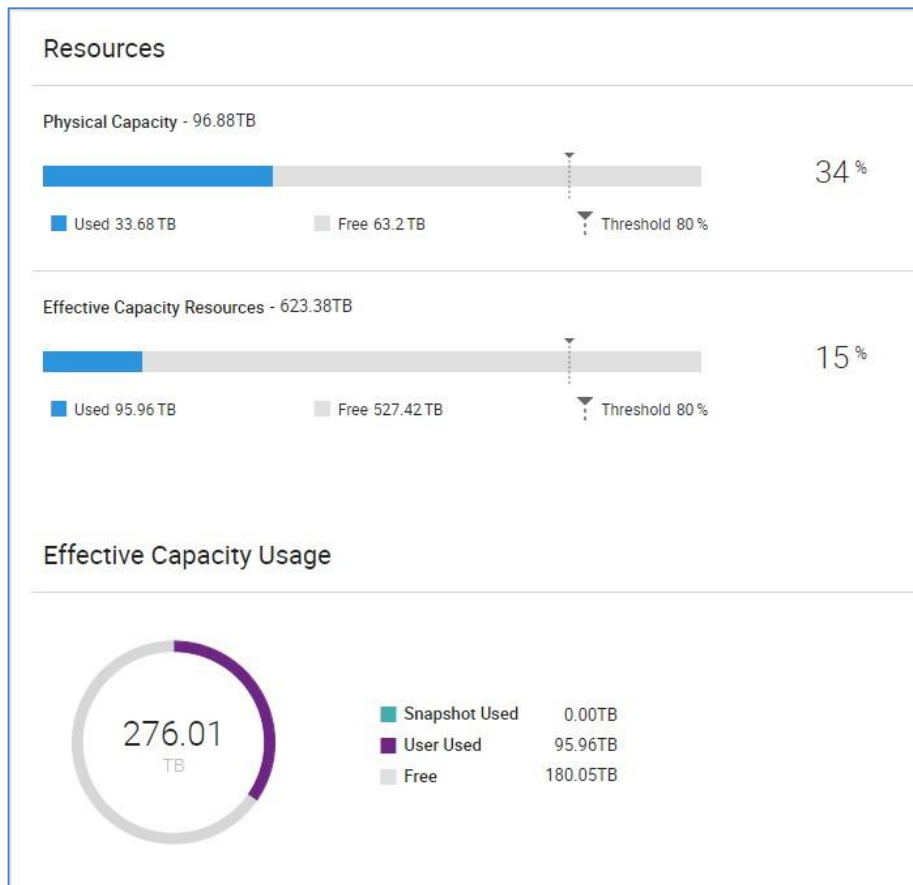


Figure 5. Utilisation des ressources avec affichage de la capacité effective

Snapshot

La capacité des snapshots back-end peut être nettement inférieure aux deltas par snapshot en raison de l'efficacité des fonctionnalités telles que les allocations partagées et la réduction des données.

Passez le pointeur de la souris sur le graphique à barres Snapshot sur le tableau de bord Capacity pour obtenir des informations générales. Les valeurs de snapshot sont définies comme suit :

- **Used** : capacité effective utilisée pour les données de modification de snapshots
- **Free** : quantité restante de données de modification de snapshots en fonction de la quantité utilisée et des métadonnées restantes
- **Total** : Used + Free
- **Threshold** : seuil d'alertes sur les données de modification de snapshots

Cliquez sur les graphiques à barres pour accéder au tableau de bord Effective Capacity et au tableau de bord Snapshot Capacity.

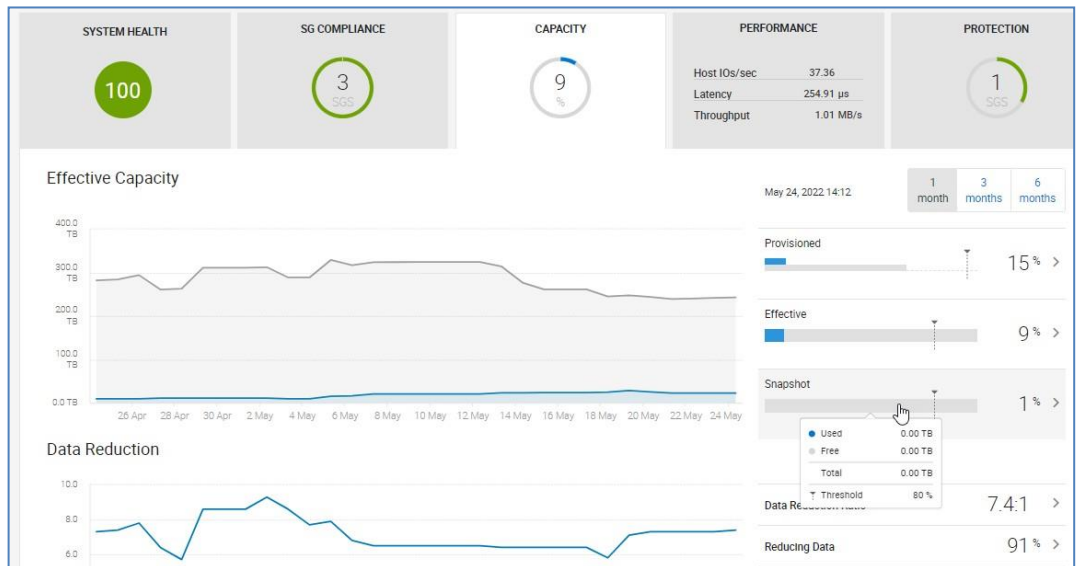


Figure 6. Tableau de bord Capacity

Contenu du tableau de bord Effective Capacity :

- **Snapshot Used** : quantité de capacité effective utilisée par les données de modification de snapshots

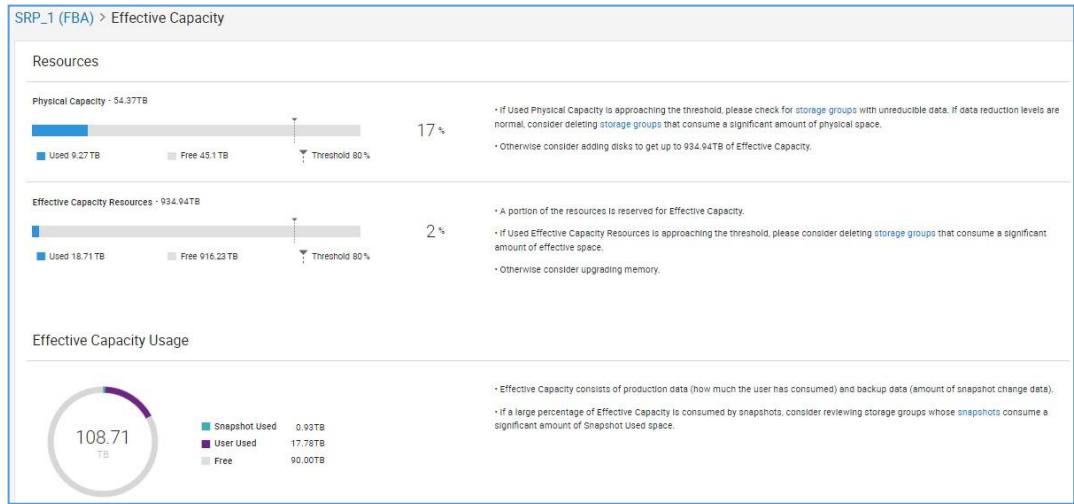


Figure 7. Tableau de bord Effective Capacity

Les valeurs de la page Snapshot Capacity sont définies comme suit :

- **Snapshot Effective Used** : pourcentage de la capacité effective utilisée par les données de snapshot pour le SRP
- **Snapshot Physical Used** : pourcentage de capacité utile consommée par les données de snapshot pour le SRP
- **Snapshot Resources** : pourcentage de métadonnées de snapshot consommées pour l'ensemble du système
- **Snapshot Resources Used** : capacité utilisée par les snapshots dans le cadre de la capacité de métadonnées de snapshots

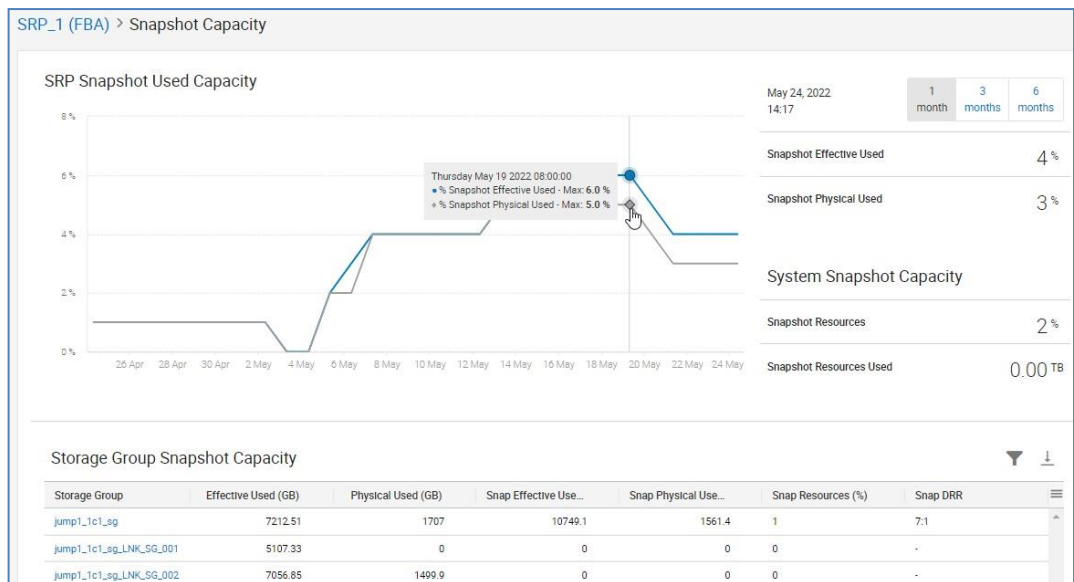


Figure 8. Utilisation de la capacité de snapshots

Taux de réduction des données

L'affichage Data Reduction fournit un emplacement unique qui permet aux utilisateurs d'afficher l'efficacité de la réduction des données. Cet écran comprend trois sections : un taux de réduction des données, un graphique interactif historique et un tableau de tous les groupes de stockage. Le taux de réduction des données indiqué prend en compte uniquement les données activées et la réduction des données écrites sur le système.

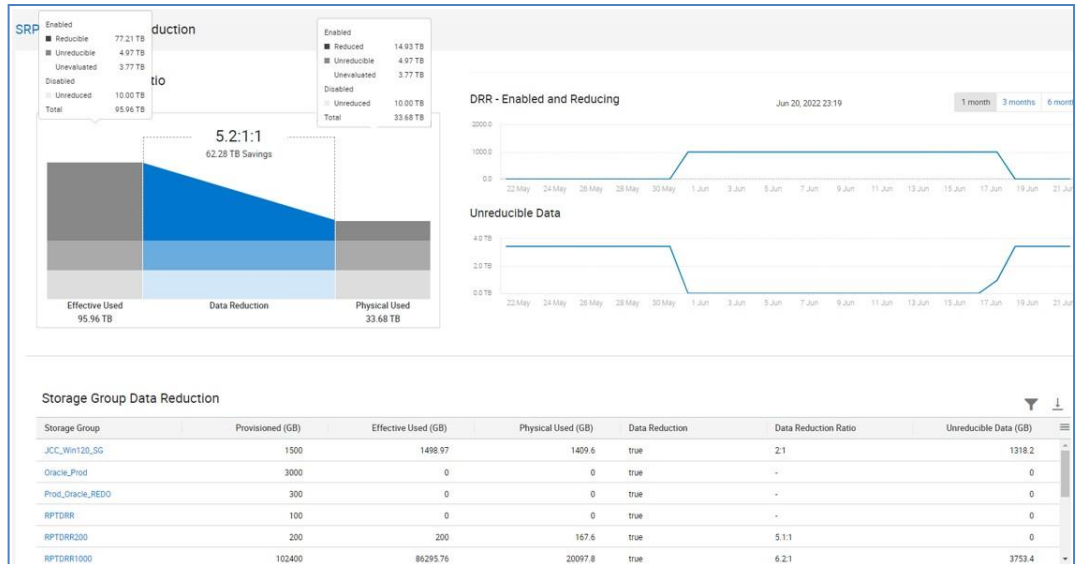


Figure 9. Affichage Data Reduction

Le **taux de réduction des données** s'affiche sous la forme d'un graphique qui présente la capacité utilisée effective, le taux de réduction des données et la capacité physique utilisée. La section **Physical Used** fait référence à la quantité réelle de capacité physique utilisée. La réduction des données présente les économies sous la forme d'un taux.

La section **Effective Used** représente les données écrites sur le système avant que les économies ne soient réalisées à l'aide de la réduction des données. Toutes les valeurs affichées représentent la taille complète telle qu'elle a été écrite par l'hôte ou l'application. Les données écrites sont placées dans deux catégories : Enabled et Disabled.

- **Enabled** indique que la réduction des données est activée et que les données prises en compte sont soumises au processus de réduction des données et à la fonction de réduction basée sur l'activité. Lorsque la réduction des données est activée, les données peuvent être classées dans trois autres catégories : Reducible, Unreducible et Unevaluated.
 - La catégorie **Reducible** correspond à la quantité de données écrites que le processus de réduction des données a identifiées comme étant des données qui peuvent être réduites pour utiliser moins de capacité physique que celles qui ont été écrites sur le système.
 - La catégorie **Unreducible** représente les données qui ne peuvent pas être réduites.
 - La catégorie **Unevaluated** correspond aux données qui n'ont pas encore été évaluées par le processus de réduction des données. Le système n'a pas encore déterminé si les données sont réductibles ou non.
- **Disabled** indique que les données écrites sur le système ne sont pas soumises à des économies de réduction des données. Toutes les données identifiées comme Disabled s'affichent à l'état Unreduced.

Physical Used représente les données écrites sur le système une fois qu'elles ont été stockées sur le disque. Cela prend en compte toutes les données activées et désactivées, ainsi que toutes les données réduites et non réduites. Il existe deux catégories qui représentent les données stockées sur le disque : Enabled et Disabled.

- **Enabled** indique que les données ont suivi le processus de réduction des données. Ces données sont classées en trois sous-catégories : Reduced, Unreducible et Unevaluated.
 - **Reduced** indique les données qui ont été envoyées via le processus de réduction des données. Ce processus inclut à la fois le transfert via le matériel de réduction des données et le stockage sur disque. Les données réduites stockées sur le disque consomment moins d'espace disque que celles qui ont été écrites par l'hôte ou l'application.
 - **Unreducible** indique que les données ont été envoyées via le processus de réduction des données, y compris le matériel de réduction des données, mais qu'elles n'ont pas pu être réduites. Certaines données irréductibles comptabilisées dans la section Physical Used peuvent contribuer aux économies liées à la réduction des données en tant que données partagées en raison de la déduplication.
 - La catégorie **Unevaluated** correspond aux données qui n'ont pas encore été évaluées par le processus de réduction des données. Le système n'a donc pas encore déterminé si les données sont réductibles ou non.
- **Disabled** indique que les données écrites sur le système ne sont pas soumises à des économies de réduction des données. Toutes les données identifiées comme Disabled s'affichent à l'état Unreduced.

Le graphique interactif indiquant **DRR Enabled and Reducing et Unreducible Data** fournit des données historiques. Il montre l'effet des données irréductibles sur le taux de réduction des données. Ce graphique permet à l'utilisateur de suivre et surveiller les modifications du taux de réduction des données qui peuvent être causées par des données irréductibles.

La liste des groupes de stockage fournit des informations sur l'utilisation de la capacité et la réduction des données propres à chaque groupe de stockage du système. Lorsque vous utilisez le graphique interactif pour suivre les modifications apportées au taux de réduction des données, vous pouvez utiliser la liste des groupes de stockage pour identifier les groupes de stockage qui comportent de grandes quantités de données irréductibles ayant un impact sur le taux de réduction des données.

Calculating Efficiency Ratios : les données nécessaires pour calculer le taux de réduction des données sont disponibles dans une fenêtre contextuelle dans le graphique Data Reduction.

- **Data Reduction Ratio** : le taux de réduction des données est calculé à l'aide des valeurs Enabled et Reducible qui figurent sous Effective Used, et à partir des valeurs Reduced qui figurent sous Physical Used.

$$\text{Enabled Reducible} \div \text{Enabled Reduced}$$

- **Overall Data Reduction Ratio** : le taux global de réduction des données du système est calculé à l'aide des valeurs totales indiquées dans Effective Used et Physical Used.

$$\text{Total Effective Used} \div \text{Total Physical Used}$$

Services de données pris en charge

Présentation

La réduction des données est prise en charge pour le stockage FBA. Les systèmes FBA/CKD mixtes sont pris en charge dans le(s) même(s) pool(s) de ressources de stockage, mais la réduction des données pour l'émulation CKD ne tire parti que de la compression et de la réduction basée sur l'activité. Tous les autres services de données proposés dans les systèmes PowerMax et VMAX All Flash sont pris en charge. Ces services comprennent la réplication locale (SnapVX), la réplication à distance (SRDF), D@RE et VMware vSphere Virtual Volumes (vVols).

Réplication locale (SnapVX)

Les snapshots SnapVX protègent les applications sans utiliser de volumes cibles pour capturer les données de modification, appelées deltas. Les deltas de snapshots sont automatiquement conservés au niveau du back-end de stockage à l'aide de pointeurs vers les images pertinentes à un point dans le temps. Le partage des ressources et la déduplication des données

Conclusion

tirent automatiquement parti de cette conception pour offrir des avantages en matière de cache, de capacité et de performances.

Les données sources compressées restent compressées lorsqu'elles deviennent des deltas de snapshots. Les données sources non compressées peuvent être compressées au fur et à mesure qu'elles deviennent des deltas de snapshots ou qu'elles deviennent moins actives. L'activité de lecture via les cibles liées peut empêcher la compression d'un delta non compressé ou entraîner la décompression des deltas de snapshots compressés. Les deltas de snapshots sont disponibles pour la déduplication.

L'activation de la réduction des données sur une cible liée n'aura d'effet que sur les données détenues par la cible liée. Les données sur les cibles et les clones liés sont disponibles pour la déduplication.

Réplication à distance (SRDF)

La compression pour SRDF est prise en charge et connue sous le nom de compression SRDF. La compression SRDF est une fonctionnalité conçue pour réduire la consommation de bande passante lors de l'envoi de données vers et depuis les systèmes connectés via la réplication à distance. La compression SRDF et la réduction des données utilisent toutes deux le même matériel, mais elles ont des objectifs différents. Les données qui ont été compressées à l'aide de la réduction des données ne sont pas compressées avant d'être transmises via la liaison SRDF. Si la compression SRDF et la compression à la volée s'appliquent, les données sont décompressées, puis compressées à l'aide de la fonction de compression SRDF, puis transmises au site distant.

Chiffrement des données au repos (D@RE)

D@RE fournit un chiffrement back-end matériel sur baie, et la réduction des données est effectuée sous la forme d'un processus à la volée. Les données sont transmises via le matériel de réduction des données avant d'être transmises via le matériel de chiffrement. Par conséquent, les données sont compressées et/ou dédupliquées avant d'être chiffrées. Sur D@RE, les données système chiffrées sur le disque ont déjà été compressées, dédupliquées ou les deux.

Virtual Volumes

La réduction des données est prise en charge pour l'allocation des données à vVols et suit les mêmes entrées/sorties que toutes les autres données. Les entrées/sorties sont visibles sur la [Figure 1](#). La réduction des données est activée au niveau des ressources de stockage dans un conteneur de stockage vVol, car il n'y a pas de groupes de stockage pour les vVols.

Conclusion

Résumé

Dans le secteur du stockage, l'utilisation de la capacité de stockage physique est une préoccupation commune des administrateurs de stockage. Les volumes de données constants et toujours croissants ont créé le besoin d'accroître l'efficacité de l'utilisation de la capacité physique. Les systèmes de stockage de données Dell PowerMax 2500 et 8500 optimisent cette efficacité. La réduction des données permet d'économiser énormément de capacité tout en offrant des performances optimales. Cela permet de réduire l'empreinte du datacenter et de diminuer globalement le coût TCO. En plus des économies réalisées, l'utilisation de la réduction des données s'active ou se désactive aussi facilement qu'un simple clic. Le système gère tout le travail.

Références

Documentation Dell Technologies

La documentation Dell Technologies suivante fournit d'autres informations relatives à ce document. L'accès à ces documents dépend de vos informations d'identification. Si vous ne parvenez pas à accéder à un document, contactez votre responsable de compte Dell Technologies.

- [InfoHub sur PowerMax et VMAX](#)