DATACORE | Swarm

Protection des données à l'ère du Cloud grâce au stockage objets DataCore Swarm

par Don Baker, ingénieur principal chez DataCore

Face à l'avalanche de données que connaissent les organisations à l'ère du Cloud, il n'a jamais été aussi important de protéger toutes ces données. Les rôles sans doute les plus évidents de Swarm consistent à stocker vos données et à les rendre accessibles, mais Swarm permet également de protéger vos données. Pour utiliser une métaphore, Swarm ressemble à un navire qui transporte et protège vos données sur un fleuve de matériel qui évolue au fil du temps, soit pour des besoins de mise à niveau, soit à cause de défaillances éventuelles.

TOUT EST PRÉVU

Cela vaut la peine d'expliquer les différents mécanismes de protection des données de Swarm et la façon dont un cluster Swarm peut servir de protection des données unitaire, même si certaines parties du système global tombent en panne. Swarm a tout prévu pour protéger les données. Avant même qu'une panne ne se produise, Swarm protège activement vos données. Notre stratégie de base consiste à faire plusieurs répliques de vos données afin de ne jamais mettre « tous nos œufs dans le même panier ». Le processeur d'intégrité de Swarm a pour fonction centrale de maintenir le bon nombre de répliques de tous les objets du cluster à l'emplacement approprié, et ce quelle que soit l'évolution des conditions. Le nombre de répliques que Swarm stocke pour un objet est déterminé par l'administrateur. Plus il y a de répliques et plus la protection est élevée, mais au détriment de l'espace qui est davantage utilisé.

Beaucoup de clients choisissent trois répliques, ce qui protège toutes les données d'un cluster contre 2 pertes de disque simultanées au prix de 3 fois plus de données logiques stockées. Toutes les répliques sont parfaitement équivalentes, donc il n'y a aucun risque qu'une réplique donnée soit plus vulnérable qu'une autre. Ces trois répliques sont placées dans des endroits du cluster peu susceptibles de tomber en panne en même temps. Nous nous assurons que toutes les répliques sont des copies fidèles en vérifiant les hachages calculés lors du transfert.

caractéristique brevetée de Swarm qu'aucun autre fournisseur de stockage ne propose. Nous permettons même à ces protections d'évoluer au fil du temps, pour que vous puissiez les adapter à l'évolution de la valeur de vos données.

PROTECTION DES DONNÉES EN TRANSIT

Swarm peut protéger vos données en transit sur le réseau à l'aide de sceaux d'intégrité (hachages) qui empêchent la falsification ou les erreurs de transmission. Ces sceaux sont stockés avec chaque objet et revérifiés lorsque l'objet est récupéré à partir de Swarm. Une demande d'écriture ou de mise à jour peut même nécessiter que toutes les répliques soient effectuées.

GESTION DES DÉFAILLANCES DE DISQUE

Enfin, les données de Swarm sont stockées sur des disques qui offrent une haute densité de données, une persistance peu coûteuse et des vitesses de transfert rapides. Bien que ce soit rare, les disques peuvent perdre des données en raison de secteurs défectueux. Ces secteurs défectueux peuvent contenir un ou plusieurs de vos objets, qui sont tous illisibles. Le processeur d'intégrité de Swarm lit régulièrement chaque objet sur le disque pour vérifier l'intégrité de ses données. Si cette vérification échoue, la réplique est déclarée non valide. Elle est remplacée par une nouvelle réplique effectuée à l'aide des données en double qui résident ailleurs dans le cluster. Nous comptons sur le fait que nous vérifions les données beaucoup plus rapidement que le taux de défaillance, pour que ce type d'erreurs de lecture/écriture de disque n'entraîne pas de perte de données..

Le plus souvent pourtant, des disques tombent en panne. Ils peuvent se dégrader lentement au fil du temps ou rapidement et de manière catastrophique. Pour les défaillances lentes, Swarm surveille l'accumulation des erreurs de disque, qui est en général le signe d'une défaillance imminente. Lorsqu'il est déterminé qu'un disque est proche de la panne, nous le « retirons ». Le retrait implique que le cluster fasse une récupération active du disque douteux sans augmenter la charge sur celui-ci. Comme il existe assez de répliques ailleurs dans le cluster, le disque retiré peut effacer ses répliques, et le résultat est un disque vide dans le cluster que l'administrateur peut remplacer à sa guise. Tout cela se fait automatiquement et sans aucun risque de perte de données.