

Comprendre les NAS d'entreprise

Compte rendu de la présentation du 10 septembre 2013 – Pavillon Dauphine

Intervenant M. Peter LONG - SNIA

Compte rendu rédigé par Isabelle Mauranges & l'ANDSI

En bref...

Aujourd'hui, la croissance continue des données de l'entreprise la conduit à recourir aux solutions techniques dites NAS (Network Attached Storage) d'entreprise. Ces solutions de stockages, évolutives en capacité et en performance, autorisent la connexion de plusieurs milliers d'utilisateurs simultanés. Peter Long présente l'état de l'art de ces solutions et, en particulier, les différences entre scale-up et scale-out.

Fort de son expérience et tout en défendant ces solutions, il attire notre attention sur la démarche méthodologique à respecter avant de se plonger dans le NAS : plusieurs questions doivent être préalablement posées : de quelles performances ai-je besoin (IOPS) ? ai-je l'usage d'un cache ? quelle est la taille moyenne de mes fichiers et la nature des accès séquentiels ou aléatoires. Quelle vision de l'entreprise à 5 ans ? Et qu'en est-il de mon équipe technique ? ...

Une fois brossé ce tableau d'avenir, Peter Long nous initie à de multiples notions techniques accompagnant le sujet ... Réplication, backup, sur site, hors site, quelques Go ou plusieurs Po, ainsi que tiering, caching, multi-tenancy, déduplication et compression

L'Association Nationale des Directeurs des Systèmes d'Information organise des débats et en diffuse des comptes-rendus, les idées restant de la seule responsabilité de leurs auteurs. Elle peut également diffuser les commentaires que suscitent ces documents.

Exposé

Le SNIA

Le SNIA (Storage Networking Industry Association, www.snia-europe.org) est un organisme créé aux USA par les principaux constructeurs de stockage SAN (Storage Area Network) ou NAS (Network Attached Storage). Il a pour but de faire connaître les travaux d'élaboration des standards menés au niveau international et de favoriser l'échange des connaissances.

Qu'est qu'un NAS ?

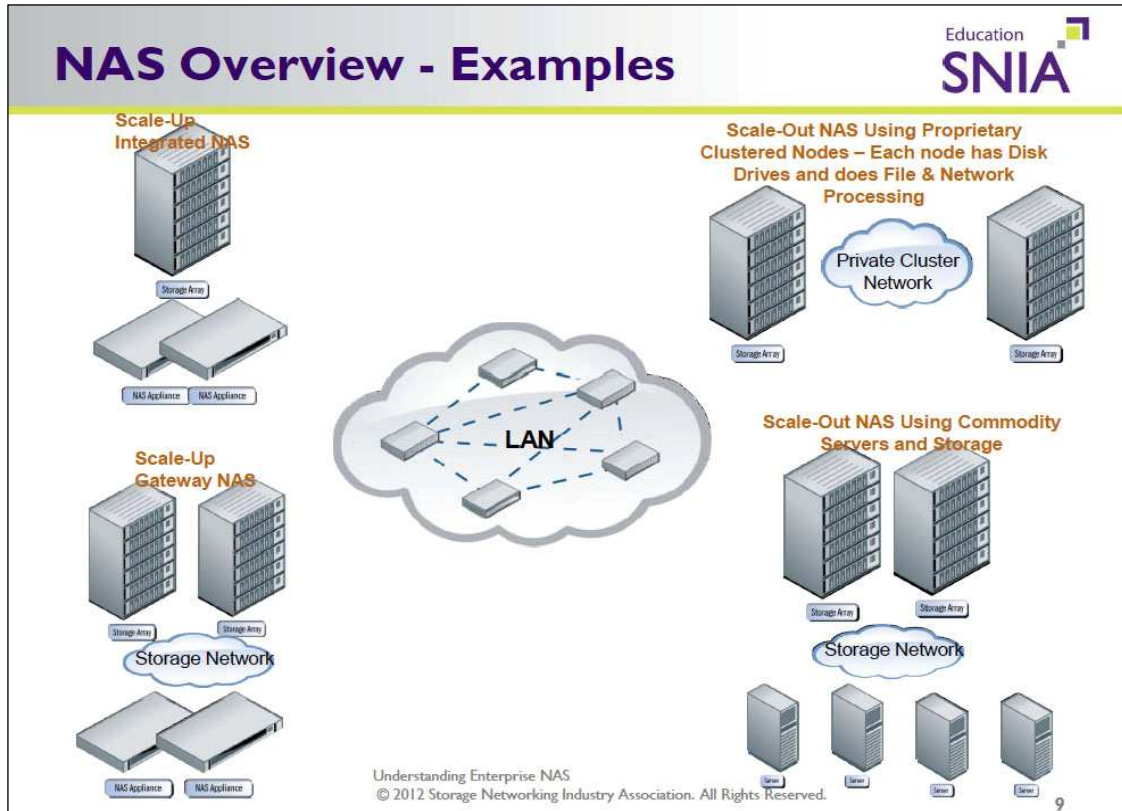
Le NAS traditionnel est une mini appliance qui permet d'accéder au travers des protocoles NFS (Network File System) ou CIFS (Common Internet File System) à un stockage basé sur du réseau IP. C'est un stockage attaché à un LAN (Local Area Network) qui va fournir de l'espace à un ou plusieurs utilisateurs. Cet espace est traditionnellement accessible via le réseau Ethernet. Ces NAS traditionnels sont polyvalents et fournissent plusieurs services au sein d'une même appliance : service mail, service impression... Une particularité du NAS est d'être plug and play, non évolutif en termes de capacité ou de performance.

A contrario, le NAS d'entreprise est évolutif en termes de capacité et de performance et il autorise des connexions simultanées importantes (de 100 à plusieurs milliers d'utilisateurs simultanés). Le NAS d'entreprise permet la redondance et la disponibilité des données (7/7 - 24/24), la hiérarchisation et le placement automatique des données (i.e. le tiering) afin de gérer le cycle de vie de la donnée, le cache pour accélérer l'accès aux données, la déduplication et le multi-tenant (par opposition au multi-instance, le NAS physique peut être découpé en tronçons réservés à des ressources particulières) et enfin l'accès multi-protocole (CIFS ou SMB - Server Message Block - et NFS).

Les différentes approches de NAS d'entreprise

L'approche Scale-up (approche verticale) permet d'accroître la capacité. Le Scale-up permet d'ajouter des disques lorsque que l'on a besoin de plus de capacité. Dans une approche Scale-up, on ne peut pas faire évoluer les têtes NAS (cf. les deux premières questions, ci-après), on ne peut qu'ajouter du disque. Les NAS d'entreprise en scale-up sont basés sur des appliances qui peuvent être propriétaires ou basées sur un système en Open source et fournir un accès SAN et NAS.

De son côté, l'approche Scale-out relève plus du logiciel propriétaire avec file system distribué. Pour évoluer on ajoute alors des appliances, ou des têtes NAS et du disque.



Sur l'approche Scale-up, à gauche, on trouve deux têtes NAS (appliances) l'une est active, l'autre passive. On peut également avoir deux têtes actives servant les accès d'un grand nombre de clients sur un seul stockage arrière de 1 à plusieurs centaines de disques. Sur l'approche Scale-up, l'évolution peut se faire en termes de gateways : ces appliances sont des NAS connectés à plusieurs stockages. Un réseau sera dédié au stockage et équipé d'une ou plusieurs baies de stockage. A droite, l'approche Scale-out est différente : le NAS appliance est directement intégré dans le stockage. Si l'on veut évoluer en capacité ou en performance, on ajoute un deuxième châssis intégrant la partie NAS et les disques.

Enfin, la dernière approche est une approche dite de massive Scale-out qui consiste à ajouter des nœuds qui sont des serveurs standard sur lesquels est installé un système d'exploitation standard pour gérer le système de fichiers distribués. Il est dès lors possible d'aller sur une ou plusieurs baies de stockage, tous les serveurs voient le stockage et peuvent y accéder.

Que l'approche soit Scale-up ou Scale-out, les têtes NAS sont directement reliées au serveur local d'entreprise et la connexion au stockage se fait via ce réseau local.

Les caractéristiques importantes du NAS d'entreprise.

Le NAS d'entreprise doit déjà être en cluster (grappe) avec deux têtes afin d'assurer un minimum de redondance en cas de panne. La taille des environnements se mesure en Petaoctets (10^{15} octets). Le support de plusieurs types de disques est un plus. En effet, cette fonctionnalité est essentielle pour les clients qui recherchent une grande performance sur certaines données, plus importantes ou plus critiques que telles autres qui peuvent être stockées sur un disque moins couteux et plus capacitif (par exemple un disque 7000 tours de type Nearline SAS).

La classification des données est une composante non négligeable : la baie ou les têtes NAS sont capables de suivre le cycle de vie de la donnée, de connaître les performances d'accès et de décider si la donnée doit être placée sur des disques moins performants. La donnée demeurera toujours accessible mais à un coût moindre et en libérant de la place :

- Notion de cache pour accélérer l'accès à la donnée ;
- Notion de déduplication, faite en mode bloc, en mode fichier, ou les deux ; La compression ;
- Notion de single name space : faire voir une entité logique quelque soit le nombre de nœuds ou de stockage qui se trouve associés ; le single name space conserve la même vue logique « mon NAS » par exemple et ceci quelle que soit l'architecture (nombre de serveur, nombre de partage, etc.), on y accède de la même manière ;
- Enfin, le multi-tenancy (cf. Question 5) permet de scinder ou d'isoler des ressources particulières pour tel département, telle unité, tel applicatif au sein d'une même entité physique.

Les pré-requis

Suivant les types de charges, les pré-requis seront différents. La performance de disques est mesurée en IOPS (Input/Output Per Second - opérations d'entrées/sorties par seconde). Ces IOPS sont généralement fournies par la configuration disques. Suivant son type, un disque pourra délivrer un certain nombre d'entrées/sorties par seconde. Par exemple, un système qui ne dispose que de quatre disques 7 000 tours ne peut pas fournir 200 000 IOPS car c'est mécaniquement impossible. Il faut donc faire attention aux objectifs de performance I/O et la plus part du temps c'est l'équipe avant vente qui aide à définir la performance.

A la différence du monde SAN, le monde NAS utilise plutôt la notion « d'opération par seconde ». En effet, lorsque l'on réalise un listing d'un directory ou un partage de disques pour l'entreprise, cela ne génère pas d'IOPS mais cela représente des opérations.

Audit de configuration - Que surveiller ?

Vérifier que pour la sauvegarde l'usage du cache est bénéfique ou non. Faites un audit d'environnement : avez-vous de petits ou de très gros fichiers, des accès séquentiels ou aléatoires ? Le document de restitution de l'audit doit préciser l'accès, le nombre de connexions, le type de fichiers, le nombre d'accès aux fichiers par jours, par mois ... Ces données sont essentielles pour bien concevoir votre configuration. Autre notion importante, la réplication et le backup. Si l'on veut répliquer le NAS d'un site A vers un site B, cela va mettre une charge supplémentaire au niveau du NAS. Pour votre backup, si vous voulez sauvegarder quelques Go ou To cela pourra se faire. Mais, en revanche, si l'on parle d'un millier de disques, de volumes en Po, cela devient vraiment très difficile, sauf à avoir une infrastructure extrêmement sophistiquée !

Tous ces paramètres sont à prendre en compte et il faut savoir, dès le départ, comment vont s'effectuer les sauvegardes, par quel protocole, quel mécanisme et doit être évaluée la charge supplémentaire que va devoir absorber le NAS.

La déduplication

La déduplication est un mécanisme qui permet d'analyser les fichiers et de détecter, en particulier, les doublons, cela au détriment de consommation CPU (Central Processing Unit, du calcul).

Les contraintes

Sur la partie scale-up, ces contraintes sont liées au niveau des disques puis des têtes. Lorsque les disques fournissent leur maximum en termes de performance et qu'il n'est pas possible d'aller plus loin, sauf à rajouter des disques supplémentaires. Ensuite, ce sont les têtes NAS qui elles aussi vont saturer, il faudra les changer en leur donnant plus de CPU ou plus de mémoires.

Une autre notion importante concerne le stockage accessible via le réseau qui doit être proportionné à la charge. Une fois rajoutés des disques et installées des têtes plus puissantes, votre réseau sera-t-il assez robuste pour supporter 10 000 clients qui peuvent se connecter simultanément ? Enfin, ce réseau doit aussi être mis à jour. (par exemple, pour accroître la bande passante de votre connexion, si vous êtes relié à un lien en Go, vous pouvez accroître la capacité du lien, ou en augmenter le nombre, les agréger ...)

Comment jauger les performances ?

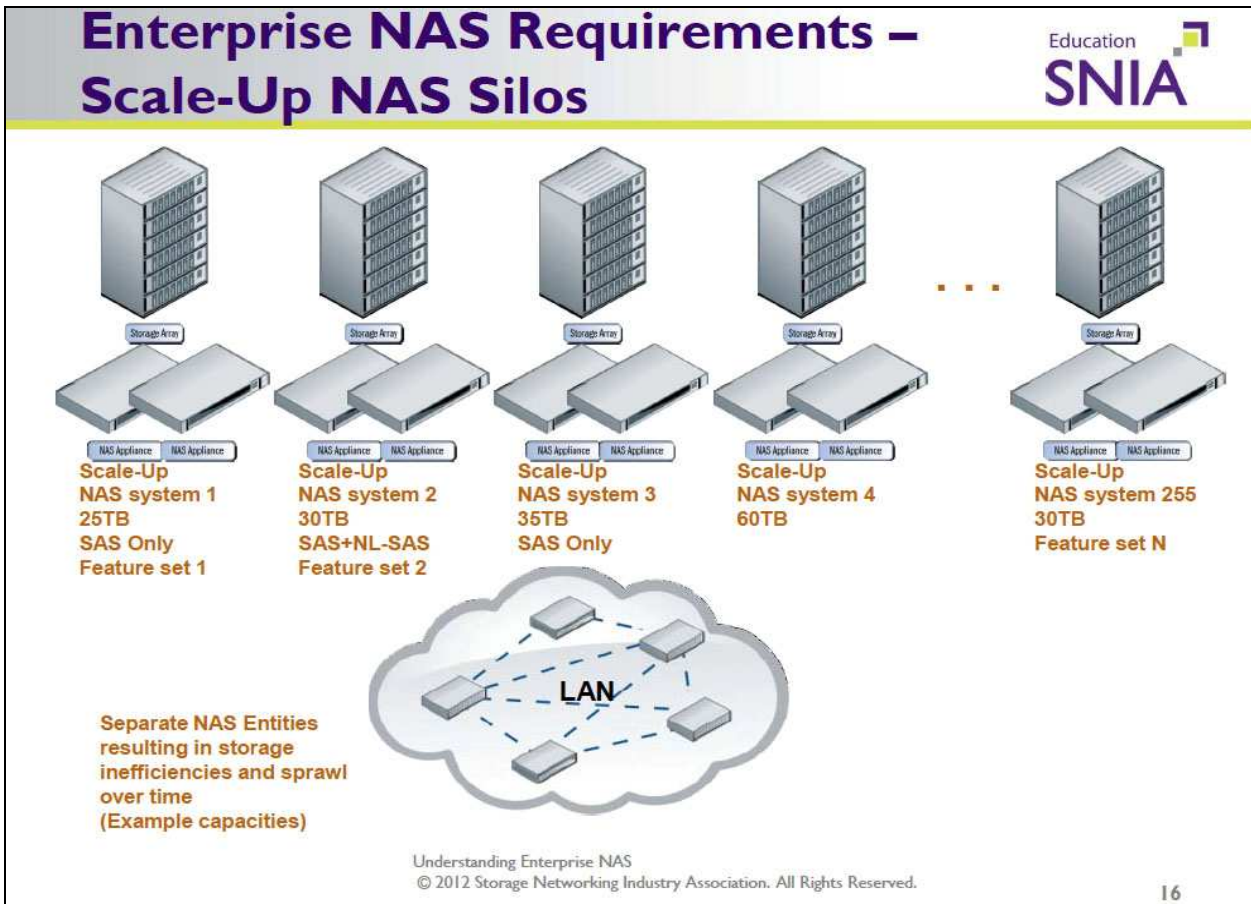
En commençant par réaliser un audit. En partant d'une infrastructure déjà existante, avec un système NAS fournissant des statistiques, il est possible de jauger les performances actuelles puis extrapoler les besoins des années futures en termes de projet et de service. En revanche s'il s'agit d'un nouveau service, il faut le parangonner (ou benchmarker) en fonction de l'activité prévue (imagerie, vidéo, la taille et le nombre de fichiers utilisés, le nombre de connexions ...) pour connaître les niveaux de performance requis Le parangonnage sera combiné avec la partie backup,

déduplication et réplication afin de connaître l'impact sur la production lorsque X clients se connecteront ou écriront Y milliers de fichiers simultanément.

Evolutivité

Il faut donc avoir une vision à long terme du taux de croissance, de la stratégie souhaitée et il faut disposer d'une vision de l'entreprise à 5 ans. Cela permet de déterminer les approches scale-up et scale-out et d'intégrer suffisamment tôt les besoins en termes de sauvegardes, d'archivage, de réplication, de déduplication, afin d'éviter le rajout intempestif de disques ou des têtes NAS. Un autre paramètre à prendre en compte est l'évolution de l'équipe technique et de votre infrastructure.

Les différentes approches



Les notions de scale-up.

Certaines entreprises préfèrent grossir par silos. Si le besoin s'en fait sentir, elles rajoutent un 2^e silo. Mais en terme d'infrastructure, l'équipe technique va avoir à gérer 1, 2, ... , 250 NAS ce qui devient vraiment important. Au lieu d'agir en termes de silos, il faut alors penser à partir sur une approche Scale-out avec un seul NAS à gérer. Ce scénario est plus rentable à gérer en terme de capacité.

Les visions à court terme.

Ajouter ou remplacer la solution existante ? Certaines entreprises préfèrent changer leurs configurations tous les ans plutôt que de rajouter des NAS ou de la capacité mais il y a risque au niveau de la migration des données qui peut entraîner un coût (caché) important.

Les caches « Frontaux »

Les caches sont des éléments que l'on insère au niveau de la tête NAS et qui permettent d'accélérer l'accès direct aux données. Les types de caches les plus communs sont : de la mémoire RAM, des cartes ou des disques SSD (Solid-State Drive, unité de stockage purement électronique).

Ces matériaux vont être considérés comme du cache et les données consultées seront copiées directement dans le cache puis re-dispatchées. Le cache peut être également bénéfique pour la phase d'écriture. Lorsque vous écrivez

des fichiers, ils seront écrits directement dans le cache puis la baie, ou le NAS, se chargera de les écrire sur les disques de manière asynchrone.

Il y a une phase d'apprentissage, d'adaptation ; Il faut donc déterminer l'adéquation de la charge de travail en termes de charges connues ou prévisibles, de données structurées ou pas, de quotas de lectures aléatoires et séquentielles. La part des données connues ou prévisibles sera systématiquement montée dans le cache. Autre notion importante : savoir si le cache est ou non volatil ! En effet, en cas d'arrêt électrique, il est essentiel que le contenu du cache soit maintenu, surtout si l'on fait de l'écriture.

Notion de tiering et caching

Sur la partie tiering (classification des données) et caching, peut-on implémenter du SSD (cf. Question 6) ? Ces deux dispositifs sont complémentaires : le cache est réservé pour accélérer l'accès aux données tandis que le tiering est plus dépendant de la gestion du temps (par exemple : si vous avez écrit une donnée, rien ne dit que dans deux ou trois jours il ne faudra pas la déplacer sur des disques moins coûteux...). Le tiering connaît aussi une phase d'apprentissage afin de déplacer intelligemment les données au sein de ce système. Le tiering peut être implémenté en mode bloc ou mode fichiers. Enfin, un tiering fait intelligemment par le système, donne moins de travail aux équipes techniques.

Ne pas oublier de prendre en compte également les backups et répliquions. Enfin, autre notion importante pour le tiering : reporting et monitoring permettront d'améliorer les process. Que se passe-t-il dans les baies ? Où sont placées les données et pourquoi ? Comment les volumes sont ils répartis sur les disques ? Comment voir la partie active et inactive au sein d'un volume.

Les notions Scale-up et Scale-out.

L'approche scale-up, ou évolution en mode vertical, correspond à l'ajout de disques dans un Système existant, lui permettant de croître en termes de capacité et de performance. Les systèmes Scale-up sont multi protocoles et ce sont des systèmes unifiés proposant des accès SAN et NAS sur le même stockage.

L'approche scale-out, évolutive en mode horizontal, correspond à l'ajout de nœuds dotés de stockage pour accroître la capacité et avoir aussi de la performance. Elle permet une meilleure utilisation des disques. Cette approche est essentiellement liée au monde de la vidéo et de la production, l'imagerie, monde pétrolier. Cette approche est basée sur un file system distribué. A noter que pour cette approche, la partie réseaux est à prendre en compte (en raison du plus grand nombre de plus de nœuds) car cela entraîne un accroissement des consommations électriques et des besoins de refroidissement ainsi qu'une emprise au sol plus importante au sein du data center.

Les pour et les contre des 2 approches

Cela dépend des besoins et des scellés que vous allez fixer. Dans l'approche Scale-up avec deux têtes NAS, si vous avez une panne sur la première tête, vous pouvez basculer sur la 2^{ème}, le service est toujours actif, alors que dans l'approche scale-out, suivant le nombre de nœuds installés, vous allez permettre 1, 2 ou X pannes. Le choix dépend avant tout des services que l'on souhaite fournir et des interruptions garanties ou non aux utilisateurs. Prendre en compte également les protocoles disponibles ou non selon les approches Scale-up ou Scale-out. Et tenir compte des équipes techniques. Si elles sont familières avec un protocole, cela ne sert à rien d'en changer.

La déduplication et la compression

La déduplication (suppression de fichiers redondants), la compression (suppression de bits redondants). L'idéal est d'avoir les deux systèmes. La déduplication n'est pas pour tout le monde car elle peut pénaliser la performance du système NAS et avoir un impact sur le temps de réponse.

EN SYNTHÈSE

Il existe beaucoup de fonctionnalités et de NAS d'entreprises avec des approches différentes Scale-up ou out. Il faut avant tout tenir compte de ses besoins et de ses stratégies. Les solutions techniques se valent mais il semble que le courant porte plutôt vers le Scale-out qui est un peu plus novateur. Tout le cloud suit l'approche Scale-out car les besoins de capacité future sont inconnus. Partir en Scale-up sur des évaluations fausses nécessite de tout changer !

Débat

Intervenant : Qu'est ce qu'une tête NAS ?

SNIA : C'est un contrôleur qui va gérer les entrées et sorties ou les requêtes des clients qui accèdent à un service/partage (ex. un lecteur réseau). Ce lecteur est émulé par la tête NAS qui montre le stockage via le réseau IP et va gérer tous les accès pour l'inscrire sur le disque.

Int. : Combien de disques sont gérés par les têtes NAS ?

SNIA : Cela dépend des constructeurs. Un NAS traditionnel c'est 2 ou 3 disques en interne. Sur les NAS d'entreprise avec une approche Scale-up c'est de 10 à 100 disques voire plus jusqu'à un millier de disques au maximum.

Int.: Quelle est la différence entre le NAS et le SAN ?

SNIA : Le SAN gère des blocs situés sur les plateaux au niveau des disques. Une baie SAN ne se soucie pas de la cohérence des blocs, qu'il y ait beaucoup de 0 ou de 1, ou que le *file system* soit corrompu, la baie SAN fournit des blocs sur lesquels vous écrivez. Le système de fichier est lié au système d'exploitation sur le serveur. Pour le NAS, le système de fichiers est géré directement par les têtes NAS. Il n'est plus géré par le serveur qui sert uniquement à communiquer avec la tête NAS, via un protocole. Cela ressemble plus à un proxy. La tête NAS ne fournit pas de blocs mais un espace de stockage avec protocole de communication.

Int.: Quelle est la différence de coût entre NAS et SAN ?

SNIA : Le coût dépend de l'infrastructure. En termes de NAS les entreprises possèdent déjà généralement l'infrastructure réseau IP et l'on peut se brancher dessus. Sur la partie SAN il y a déjà la baie à acheter et derrière il y a l'infrastructure réseau SAN à ajouter avec les switchs ce qui représente un coût supplémentaire. Le coût caché est aussi celui des compétences techniques à prendre en compte.

Int.: Quel est l'intérêt de multi-tenancy ?

SNIA : La sécurité peut être un intérêt mais surtout pour des départements spécifiques qui veulent avoir leur propre stockage, Le multi-tenancy est très utilisé pour les providers. Par exemple, vous avez un fournisseur qui vous offre du stockage, vous êtes le client A, il va vous octroyer tant de CPU, de mémoire, de réseau et de stockage. Dans une entité logique qui est la vôtre. Ils ont un autre client, le client B, il va faire la même chose et garantira que le client B ne peut pas prendre des ressources sur la partie allouée au client A.

Int.: Le SSD est-il considéré comme une catégorie particulière de tiering ?

SNIA : Vous avez des classes de tiering dans lequel on va incorporer des types de disques différents dont le SSD, mais on souhaite que les disques les plus performants soient utilisés lors des écritures. Au bout d'un certain temps, cela va être déplacé du SSD dans d'autres classes de disques...

Présentation de l'orateur

Peter LONG, spécialiste des solutions Compellent et FluidFS, a développé depuis plus de 10 ans une expertise dans le domaine du stockage SAN & NAS. Il est aujourd'hui Storage Enterprise Technologist au sein de Dell France. Intervenant dans le cadre du SNIA, dont il est membre en France, il partage ses compétences sur ces thèmes.